THE SPATIAL ORGANIZATION OF TRANSCRIPTION IN E. COLI

by

Xiaoli Weng

A dissertation submitted to Johns Hopkins University in conformity
with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland
Nov, 2017

Abstract

Transcription, the process of converting genetic information stored in DNA to RNA, lies at the heart of gene expression. Transcription has been studied extensively *in vitro* to probe its mechanistic detail; however, these conditions differ vastly from the complex environment inside a living cell. Spatial distributions of molecular components have been shown to be an important facet of gene regulation in living systems. First, to gain insight into the regulation of gene expression at the cellular level, we investigated the spatial distributions of various molecular components of transcription in *E. coli* and their physical correlation with each other using superresolution fluorescence microscopy (Chapter 2). Our results show that while dense RNAP clusters are highly colocalized with *rrn* operons and nascent rRNA transcripts during fast growth, these RNAP clusters are present independent of rRNA transcription activity, and are likely stably associated with the underlying chromosome structure. Second, we provided the first direct observation of transcription factor mediated DNA looping in live *E. coli*, and calculated *in vivo* looping frequencies in the context of different regulatory regimes (Chapter 3). Third, we initiated the development of a reconstituted CRISPR-Cas system to effectively visualize sequence specific genomic sites in live cells in a high-throughput manner. This system would allow for unprecedented insight into eukaryotic genome architecture in intact living cells; Chapter 4 details our efforts in optimizing the reconstituted CRISPR-Cas system first using *in vitro* assays.

Primary Reader: Jie Xiao

Secondary Reader: James Berger

Acknowledgements

First, I would like to thank my advisor Jie Xiao for giving me the opportunity to work in her lab. Her enthusiasm for science and unwavering optimism day-to-day is a great source of inspiration as we puzzled over interesting but often complicated results. Thank you to Christopher Bohrer, Kelsey Bettridge and Arvin Lagda for collaborating with me on the RNA Polymerase project in Chapter 2; in particular, Chris has major contributions to the data analyses in this project. The looping project in Chapter 3 was conceived and developed by Zach Hensel, and later became a collaborative effort together with Arvin. It was a memorable experience working closely together during the intense, wrap-up phase of this cool project. Thank you to Gina Wang, and Jason Lyu for working tirelessly on the more difficult *in vivo* experiments for the CRISPR project in Chapter 4. I would like to thank my thesis committee (Cynthia Wolberger, Robert Schleif and James Berger) for their invaluable insight, funding sources that make these works possible, and our program's support staff for all their help. Thank you to all members of the Xiao lab, past and present, for fun discussions about science and life; their generous technical and moral support is always greatly appreciated. I had a great time working with and alongside everyone in the Xiao lab. Finally, thank you to my family and friends, especially to my parents Jian Weng and Ping Li, for their unconditional love and endless encouragement over the years, I love you.

Contents

List of Figures

List of Tables

Chapter 1: Introduction to the Spatial Organization of Transcription in Bacteria

Adapted with permission from:
Weng, X., Xiao, J. Spatial organization of transcription in bacterial cells. **Trends in Genetics** (2014)

Location matters

Traditionally, bacterial cells have been viewed as bags of enzymes. Subcellular localization was thought unimportant because enzymes could reach their substrates fast enough through simple diffusion. For example, a protein molecule with a typical cytoplasmic diffusion constant of 8 $\mu m^2/s$ [1] will travel on average ~ 700 nm within 10 ms, which is comparable to the dimensions of a typical bacterial cell. In the past two decades, however, studies in bacterial cells have emerged to suggest the opposite — subcellular locations of biomolecules may matter for their functions [2,3].

The spatial organization of transcription in bacterial cells, here defined as the intracellular localization of various transcription components and their dynamic response to transcriptional activity (Figure 1.1A), may exhibit a level of organization that suggests possible functional compartmentalization; this may provide a new means of regulation of enzymatic activities by spatial colocalization or segregation. Many recent works have documented the subcellular localizations (Figure 1.1B) of four major components of transcription: genes, transcription factors, RNA polymerase (RNAP), and RNAs. Most of these efforts have been made possible by recent developments in innovative single-molecule, single-cell imaging techniques and high-throughput, large-scale biochemical methods. This field is currently still at the level of observations, and most evidence comes from a relatively small number of specific case studies. To move beyond these observations, we must critically analyze the evidence in the context of prior genetic and biochemical studies and consider whether the observed spatial localization pattern is pertinent to a specific biological function.

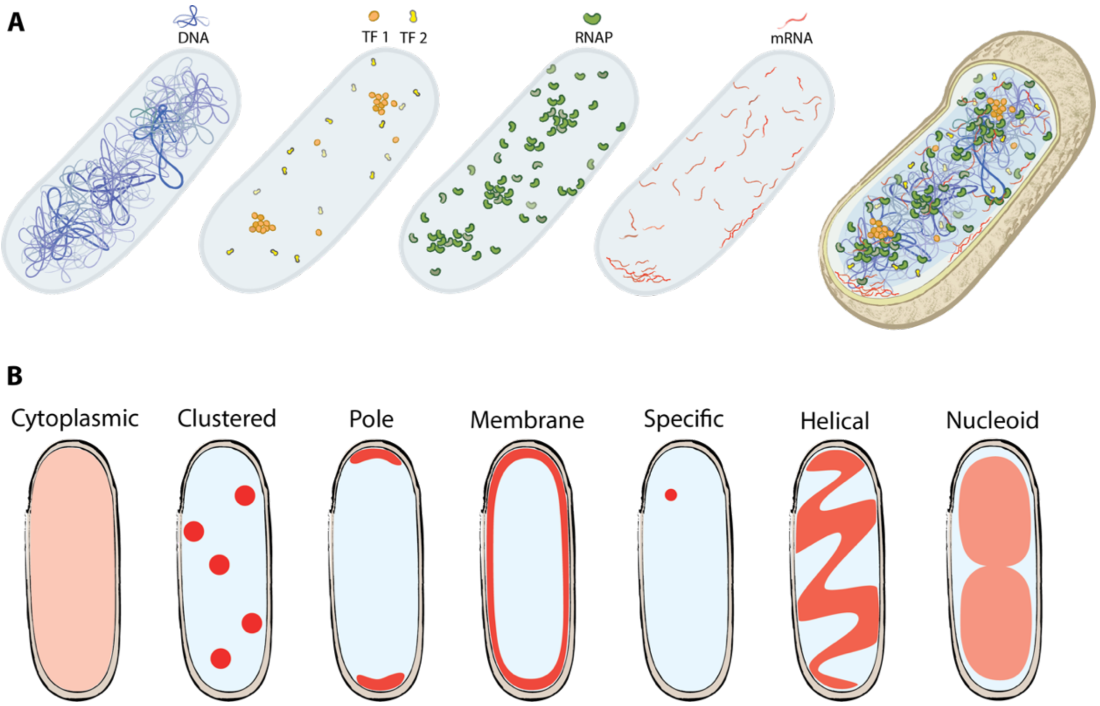


Figure 1.1 Spatial organization of transcription in an *E. coli* cell. (A) Genes (DNA), transcription factors (TF 1 and TF 2), RNA polymerase (RNAP) and mRNAs may be organized differentially in space instead of being homogenously distributed. (B) Overview of experimentally observed molecular spatial distribution patterns of transcription components in a model bacterial cell. From left to right are cytoplasmic, clustered, pole, membrane, specific, helical, and nucleoid distributions.

Organization of genes: spatial clustering versus dynamic relocating

A typical bacterial chromosome forms a compact DNA mass called the nucleoid in the center of the cell (Figure 1.2). Although there is no nuclear envelope separating the nucleoid from the cytoplasm, it is well documented that the chromosome is not a random bowl of spaghetti but is instead structurally organized

[4-9]. The typical size of a bacterial genome is a few million base pairs. This chromosomal DNA must adopt a highly compact but orderly configuration to accommodate the spatial constraints of a bacterial cell while coordinating with DNA-centric activities such as replication and transcription.

Taking the most well studied bacterial model system *E. coli* as an example, the chromosome is organized on multiple levels (Figure 1.2A). At the first level, the negatively super-coiled chromosome DNA naturally generates plectonemic loops, which are actively maintained by the opposing actions of gyrase and topoisomerases [10-12]. Protein factors such as the abundant nucleoid-associated proteins (NAPs) can influence DNA structure both locally by bending and wrapping DNA segments and globally by bridging and providing boundaries for DNA topological domains [13]. These independent topological microdomains domains are on the scale of 10-100 kb and are positioned stochastically along the chromosome [14]. The next level of organization, as assayed by both the accessibility of DNA to homologous recombination and chromosome dynamics as observed through fluorescence imaging, consists of four macrodomains of ~ 1Mbp each (*ori*, *ter*, *left* and *right*), along with two unstructured regions flanking the *ori* domain [4,5]. These macrodomains are likely further folded into a compact rod shaped nucleoid, as suggested in a recent genome conformation capture study [7]. Finally, the chromosome has a defined orientation within the cell. For *E. coli*, the replication origin and terminus are situated close to mid cell, while the left and right chromosome arms are stretched toward the two poles [15,16], this arrangement is organized and maintained by the *E. coli* SMC (structural maintenance of chromosomes) complex, MukBEF [17].

Chromosome organization in other bacterial model systems such as *C. crescentus* and *B. subtilis* show that the basic level of organization by DNA super-coiling and NAPs is similar to that of *E. coli* (Figure 1.2B, C). Furthermore, fluorescent labeling experiments have shown a helical arrangement of the chromosome in both *C. crescentus* and *B. subtilis* [8,18]. Higher resolution chromosome interaction maps obtained in recent 3C based experiments showed that the *C. crescentus* chromosome may adopt a bottle brush configuration [9]. The orientation of the chromosome in the cell is different in *B. subtilis* and *C. crescentus* compared to *E. coli.* In *B. subtilis*, the replication origin and terminus have a preference for opposite poles early in the cell cycle [18,19]. In *C. crescentus*, the chromosome is positioned with the origin always at the old pole and the terminus

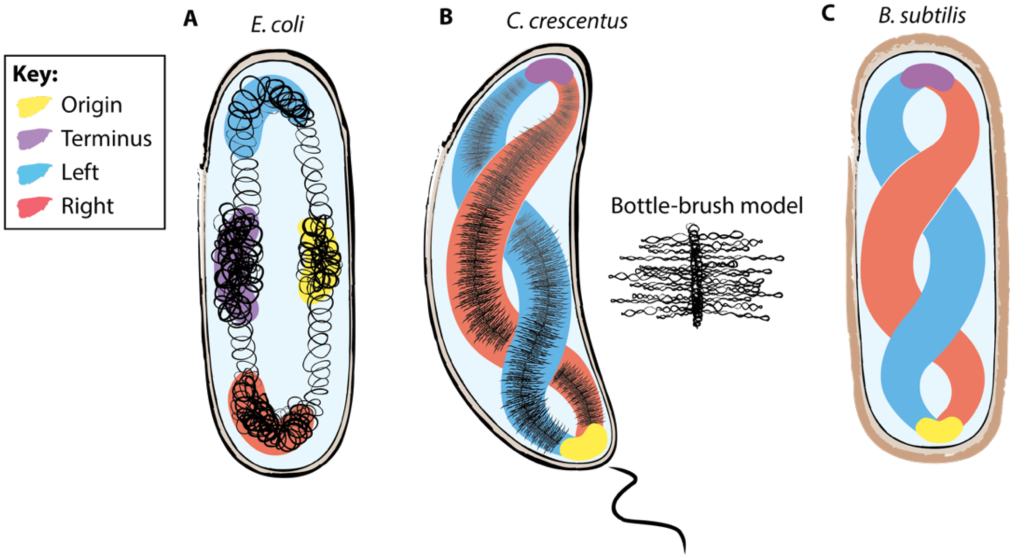


Figure 1.2 Models for nucleoid organization. Nucleoid organization in Escherichia coli (A), Caulobacter crescentus (B), and Bacillus subtilis (C). Chromosomal domains are colored as indicated by legend.

at the new pole, with the left and right chromosome arms spanning the length of the cell [8,20]. Notably, both *B. subtilis* and *C. crescentus* (but not *E. coli*) contain an origin proximal centromere like region *ParS* (Partition system), which is used for positioning of the chromosomal *oriC* region, and assists in the proper orientation of the chromosome and regulation of the cell cycle [21,22].

One of the many factors involved in organizing the nucleoid is transcription [10,23,24]. For example, nucleoids of cells treated with rifampicin, an antibiotic that traps RNAP on promoters by binding to the *β* subunit of RNAP, showed clear expansion [25-27]. Recent studies using chromosome conformation capture (3C)-based methods (Table 1.1) also found that specific chromosomal domains are established and maintained by highly expressed genes [7,9].

Table 1.1 Summary of microscopy and biochemical methods used in detecting the spatial organization of transcription.

Name	Target	Detection principle	Strength	Limitation	Ref
3C (Chromosome Conformation Capture) and derivatives	DNA	Fixation and cross-linking of genome interactions; detection by PCR, pull-down, and sequencing, etc.	Capturing native interactions in one reaction; High-throughput, large-scale detection of long range genome interactions	Interaction frequencies inferred from cross-linking efficiency Low resolution (>10kb)	28
FROS (Fluorescent Reporter Operator System)	DNA	Tandem arrays of DNA binding sites are inserted chromosomally; Detection is by the binding of fluorescent protein fused to DNA-binding proteins	- Live cell compatible - Strong signal/noise ratio if hundreds of binding sites are used - Enables the tracking of chromosome positions in real time - Orthogonal systems available: lacO-LacI; tetO-TetR; pars-ParB	Tight binding of fusion proteins on DNA is detrimental to cell physiology	29
RNA stem-loop motif – phage protein system	RNA	Tandem arrays of RNA stem-loops are inserted into genes of interest. Detection is by the binding of fluorescent protein fused to phage coat proteins that bind to the stem-loops	- Live cell compatible - Strong signal/noise ratio if hundreds of binding sites are used - Enables the tracking of real-time RNA production. - Orthogonal systems available: MS2-MCP; PP7-PCP; BoxB-lambdaN	Tight binding of coat proteins on RNA stem-loops may alter RNA stability or inhibit translation	30-33

smFISH (single molecule Fluorescence *in situ* Hybridization)	RNA	Multiple complementary short DNA oligonucleotides labeled with organic fluorophores hybridize to RNA of interest	- Detecting native RNAs - Able to achieve single RNA molecule detection	Requires fixation Hybridization efficiency varies depending on sequence	34,35
PALM (Photo-Activated Localization Microscopy)	Protein	Photoactivatable fluorescent protein fused to protein of interest to detect protein or nucleic acids. Stochastic photo-activation of single molecules allows sub-diffraction limited precision in determining the position of molecules	- Live cell compatible - 15-30 nm spatial resolution achievable	Functionality of Fluorescent protein fusions need to be verified	36
STORM (Stochastic Optical Reconstruction Microscopy)	Protein and nucleic acids	Antibodies and or oligonucleotides labeled with photo-switchable fluorophores to detect protein or nucleic acids. Blinking of dye molecules allows sub-diffraction limited precision in determining the position of molecules	15-30 nm spatial resolution achievable	Usually requires fixation	37
SMT (Single-molecule tracking)	Protein	Spatial positions of fluorescently tagged single molecules are tracked at various time points	Enables the measurement of diffusion behavior and binding/unbinding kinetics	Fluorophore bleaching and photo-blinking limits the length of tracking trajectories	38

The coupling of nucleoid structure and transcriptional activity suggests that genes may be spatially organized according to their transcriptional activities, irrespective of their linear orders on the chromosome. Intuitively, there could be two ways to organize genes spatially. The first is that genes sharing similar regulatory controls could spatially cluster together (Figure 1.3A); the second is that the cellular location of a gene could dynamically correlate with its transcriptional activity (Figure 1.3B).

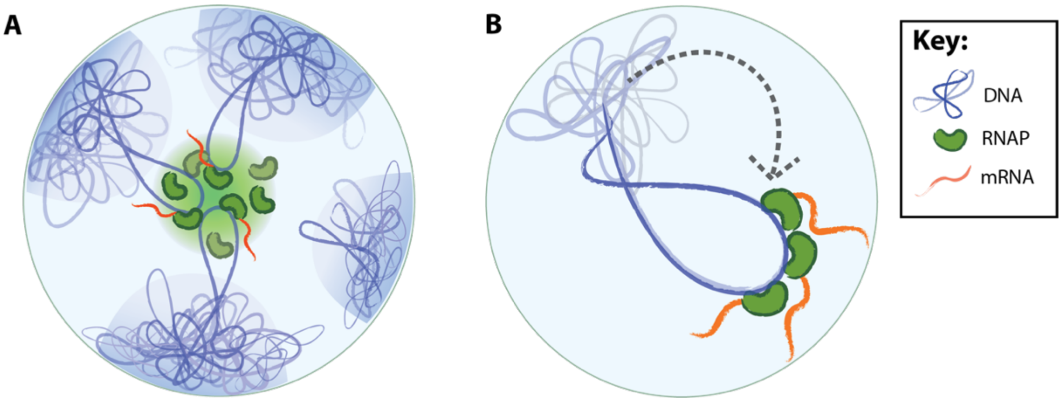


Figure 1.3 Spatial organization of genes. (A) Multiple genes spatially cluster together to be co-regulated. (B) A gene may move out of its ordinary boundary to be activated.

Spatial clustering of genes is supported by computational analyses that showed that pairs of distant genes (>100 kb) have correlated expression levels, suggesting that these genes may share a similar environment [39-43]. Spatially clustered genes may also have the advantage of confining transcription to local areas where high concentrations of RNAP and transcription factors allow rapid response and efficient transcription [44]. This scenario is similar to the transcription factory theory proposed for eukaryotic cells [45].

The transcription factory theory was initially proposed for RNAPII in eukaryotic cells [45]. In this model, multiple active RNAPII molecules form foci or clusters to reel genes through for transcription. One of the features of this model is that genes far away from each other could be brought into the same factory, and thus could be coordinately regulated by the same factors despite their physical distance (reviewed in [46]). Past studies using a combination of conventional fluorescence microscopy, EM, and 3C techniques in eukaryotic cells have well established that some active genes indeed colocalize with active RNAPII foci [47-49].

Nonetheless, many aspects of the model, such as whether RNAPII factories are stable or dynamic cellular structures, and whether it is a general mode of transcription, remain unclear.

Recently, two single-molecule based superresolution imaging studies investigated the spatial distribution of RNAPII in the same type of human cell [50,51]. One study found that RNAPII indeed forms large clusters in live cells; these clusters are dynamic, transitory, and respond to changes in the transcriptional state of the cell [50]. However, due to the live-cell nature of the study, it was difficult to quantify on average how many RNAPII molecules are in a cluster, and how many of these clusters are in a cell. The other study, conducted in fixed cells where the number of detected molecules could be accurately counted, found that the majority of RNAPII foci (>70%) only contained a single RNAPII molecule, and <10% of foci contained >4 RNAPII molecules [51]. Regardless of differences in the experimental and quantification methods used in the two studies (live vs. fixed cells, wide-filed vs. light-sheet illumination, FP fusion vs. affinity tag labeling, etc.), both studies indicate that while transcription factories may exist, they are not static cellular structures as previously thought, and may not be a prevalent mode of transcription in eukaryotic cells.

The observation of dense RNAP foci in bacterial cells inspired a similar theory in prokaryotes. However, there are large differences in eukaryotic and prokaryotic transcription. For example, RNAPI in eukaryotes is responsible for synthesizing rRNA, and localizes to the nucleolus, whereas a single RNAP is responsible for all transcription in bacteria. Furthermore, transcription foci in bacterial cells are sensitive to growth conditions and cell physiology — punctate RNAP clusters are only observed in *E. coli* cells under fast growth [52]. To date, the transcription factory theory remains controversial in both eukaryotes and prokaryotes.

The hypothesis of spatially clustered genes initially stemmed from the observation that a functional RNAP-green fluorescent protein (GFP) fusion (labeled on the β' subunit, RpoC-GFP) forms one or two dense foci per chromosome in *E. coli* cells growing in rich media (see more discussions in the RNAP section) [52-54]. Because ribosomal RNA (rRNA) synthesis is the major transcription activity in cells with high growth rates [55-57], it has been assumed that these RNAP foci are active RNAP molecules engaged in rRNA synthesis. As multiple rRNA operons are spaced far away from each other on the chromosome (seven in *E. coli* and ten in *B. subtilis* [58,59]), and the copy number of *rrn* operons in a fast-growing *E. coli* cell can reach up to 50, the observation of far fewer RNAP foci suggests that multiple rRNA operons are transcribed while clustered together [52-54].

A recent study where the spatial distribution of H-NS, one of the nucleoid-associated proteins (NAPs), was investigated in *E. coli* showed additional evidence for clustering of genes. Using single-molecule based superresolution imaging (Table 1.1), it was found that H-NS forms on average two clusters per chromosome in cells (Figure 1.4A) [60]. More interestingly, two-color colocalization showed that these clusters colocalized with genes that H-NS regulates (Figure 1.4A) [60]. In another study, the subcellular localizations of multiple *gal* operons (all regulated by the Gal Repressor GalR) were investigated using both fluorescence microscopy and 3C in *E. coli* [44]. It was found that yellow fluorescent protein (YFP)-labeled GalR molecules formed one to three punctate foci in cells at stationary phase, and subsequent 3C experiments detected interactions between multiple *gal* operons that are hundreds of kilobases away on the chromosome. In cells lacking GalR, however, such interactions were abolished, suggesting that the clustering of these operons is related to the binding of GalR.

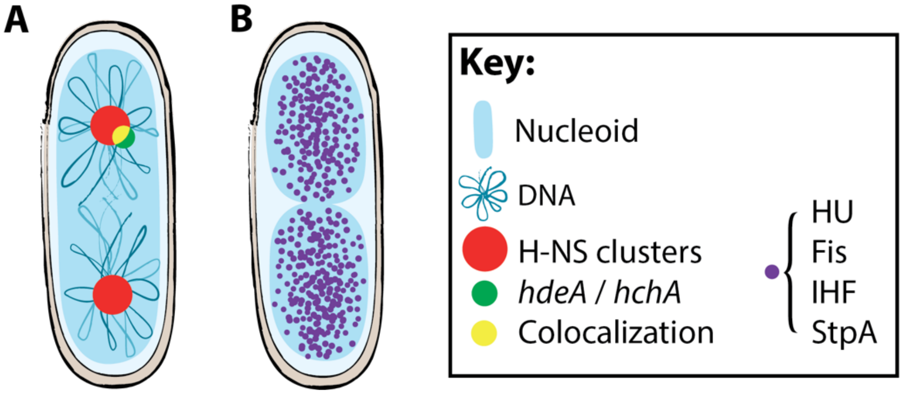


Figure 1.4 Different spatial distributions of transcription factors. In *E. coli*, global nucleoid-associated proteins (NAPs) are differentially distributed, with H-NS forming on average two clusters per chromosome (A), HU, Fis, IHF, and StpA all distributing relatively homogonously on the nucleoid (B). H-NS clusters were shown to colocalize with H-NS regulated genes *hdeA* and *hchA*; H-NS binding sites across the chromosome are inferred to be co-localized with H-NS clusters (A).

It is important to note that the studies described above used the spatial clustering of RNAP or transcription factors to infer the spatial distribution of the genes they bind to. The spatial proximity of these genes was then investigated using 3C and its derivatives, or inferred by correlating with known transcription activities under the same growth conditions (see more discussions in the section of transcription factors). The actual cellular localizations or spatial clustering of these genes were not directly visualized and compared to the protein clusters. As such, the results are still controversial. For example, it has not been experimentally proven that multiple rRNA operons colocalize with each other in RNAP foci. In fact, interactions between rRNA operons were not detected in a recent 3C study, which was attributed to technical limitations in the 3C analysis of repetitive gene loci [7]. The same 3C study also failed to detect interactions between H-NS regulated genes. This discrepancy was attributed to the improved resolution in the new 3C study. However, other experimental differences such as cell growth conditions may also contribute to the discrepancies. In addition, because 3C and its derivatives detect the juxtaposition of DNA sites by evaluating their cross linking frequencies with averaging across an ensemble of cells, transient interactions between genes may not be easily detectable [28,61,62]. As we describe at the end of this section, new methods that enable the direct, sensitive detection of individual gene locus in single cells are key to resolving these issues.

The second way to organize genes spatially is by dynamically changing the cellular location of a gene depending on its transcription activity (Figure 1.3B). The regulation of transcription activity may involve moving genes in or out of particular subcellular locations, in addition to the binding and unbinding of transcription factors in the traditional view of gene regulation. This notion may have stemmed from early electron microscopy (EM) studies, where it was shown that transcriptionally active genes and RNAP were primarily located on the nucleoid surface instead of the interior [63,64]. Intuitively, it is easier for RNAP and transcription factors to access genes on the surface than in the interior of a packed nucleoid.

The dynamic relocation of genes in response to their transcription activities has been demonstrated in eukaryotic cells [47-49,65]. However, in bacteria, direct evidence is scarce. In one *E. coli* imaging study, a plasmid DNA was labeled using the fluorescence repressor-operator system (FROS, Table 1.1, [29]). It was found that when the plasmid carried a constitutive promoter, it formed a clear focus at the edge of the nucleoid toward the cell pole [66]. In the absence of the promoter, however, the plasmid diffused randomly inside the cytoplasm. Although the plasmid DNA is not integrated onto the chromosome, the study did show that transcription activity can influence the subcellular location of a gene. In another *E. coli* study, two genes encoding different membrane proteins were found to move toward the membrane upon induction. However, this movement is most likely due to co-transcriptional translation and insertion of proteins into the membrane (transertion), so may not be a direct effect of transcription activity [67].

Other studies also used the FROS method to label specific DNA segments on the chromosome and imaged their cellular localizations in live bacteria cells. Although not specifically looking for the correlation with transcription activities, these studies found that the intracellular localization of a DNA segment is largely dependent on where it is on the chromosome and which stage of replication it is in, in contrast to the view of a dynamically relocating gene [15,19,68-74]. Furthermore, a clever experiment in *C. crescentus* found that when the chromosome is rotated by moving the chromosome anchoring sequence *parS* to a new genomic location, the global gene expression profile is not significantly altered [8]. This study suggests that gene expression activity may not be related to specific subcellular positions.

How can these seemingly contradictory results be reconciled? It is certainly possible that the dynamic relocation of genes in response to transcription activity is not necessary; the interior of nucleoid may be equally accessible to transcription as the surface. An imaging study in mouse cells found that active transcription could also occur deep inside chromosomal territories [75]. In this case, it could be that the intracellular location of a gene does not change, but the local environment of the gene changes to promote active transcription.

Alternatively, it is also possible that transcription activity-induced change in a gene's subcellular location may not be significant enough to be detected by FROS-based methods (Table 1.1). FROS traditionally uses fluorescent proteins fused to DNA-binding proteins bound to tandem arrays of hundreds of binding sites spanning several kilobases; thus, the accuracy and precision in determining the intracellular position of a gene similar in length are compromised. Furthermore, it is especially challenging when probing small changes in positions with respect to the short axis of the cell in rod shaped bacteria, which is on the scale of 1 μm.

To resolve these issues, a method allowing for direct visualization of subcellular positions of genes with high resolution should be used. A recent single-molecule imaging study used a modified FROS strategy to track the position and movement of a 2.3-kb DNA segment in live *E. coli* cells with an accuracy of approximately 40 nm [76]. Instead of hundreds of tandem arrays, the use of only three DNA binding sites spanning less than 100 bp results in a diffraction-limited fluorescent spot. The position of the binding sites can thus be determined with sub-diffraction-limited precision by fitting its fluorescence profile to a Gaussian function [77]. This method, if combined with other methods that directly visualize the transcription activity of individual genes in live cells (see below), will allow in-depth examination of whether and how genes are spatially organized in response to their transcription activities.

Spatial distributions of transcription factors: clustering or randomly diffusing?

In the conventional view, a transcription factor (TF) molecule randomly diffuses inside a cell until it encounters a specific binding site, at which it associates tightly to regulate the corresponding gene. This notion is supported by the measured diffusion constant of ~ 0.4 $\mu m^2/s$ on a YFP-labeled Lac repressor LacI [78]. As such, if the number of TF molecules is significantly greater than its binding sites, TF distribution should be fairly homogenous over the nucleoid.

A recent single-molecule study suggests the opposite, however. In this work, the spatial distribution of a Venus-labeled LacI fusion protein in an *E. coli* strain lacking a LacI binding site was imaged. It was found that LacI-Venus preferentially localizes to regions close to where its coding gene resides [79]. This biased spatial distribution may explain previous observations that the closer the target gene is to the TF-encoding gene, the stronger the regulation by the TF is [80,81]. It is interesting to note that to explain this spatial distribution, a model in which the *lacI-venus* gene is transcribed and translated inside the nucleoid is needed. This is in contrast to the hypothesis that a gene moves to the surface of the nucleoid for transcription.

When TF molecules are bound to multiple binding sites across the chromosome, the spatial distribution of the TF should reflect that of the bound DNA sites. As we have described in the first section, GalR and H-NS both form clusters [44,60], suggesting that their multiple DNA binding sites are spatially clustered as well. Another TF in *B. subtilis*, Rok, binds to AT-rich DNA and is likewise distributed non-uniformly on the nucleoid when visualized by fluorescence microscopy [82]. In contrast, other global TFs such as HU, IHF, Fis and StpA, are largely uniformly distributed on the nucleoid in *E. coli* (Figure 1.4B) [60,83].

What determines the spatial clustering of a TF? Oligomerization of a TF has been suggested as a driving force—it was found that dimerization mutants GalR^{T322R} and H-NS^{L30P} abolished the clustered appearance of the TFs and reduced the regulation of corresponding genes [44,60]. It was verified for GalR^{T322R} that the mutation did not affect the DNA-binding properties of the protein [84]. H-NS^{L30P}, however, seems to have altered DNA-binding properties [85,86], hence the abolishment of the clustering in the mutant cannot be solely attributed to the loss of oligomerization. In addition, other TFs including HU, IHF, Fis and StpA form dimers or higher-ordered oligomers [87], but do not form spatial clusters, arguing that the oligomerization of TFs may not be sufficient to bring DNA sites together. Indeed, other factors such as the growth conditions, expression levels of NAPs and the supercoiling state of the chromosome [14,87,88], may contribute to the formation of spatial clusters. For example, previously it was shown that the repression by GalR is likely mediated by a higher-order molecular complex consisting of GalR, HU and supercoiled DNA [89]. It is possible that the local chromosome structure, along with the binding of other protein factors, contributes significantly to the stabilization of the spatial distribution pattern of a TF.

What would be the functional significance of spatially clustered TFs? As discussed previously, clusters may increase the local concentration of a TF, and consequently enhance transcription regulation. In other cases, clustered TFs may bring DNA regulatory elements that are far away close to a promoter in a way that is similar to eukaryotic enhancers, and hence strengthen transcription regulation [90]. Finally, the formation of DNA loops, long and short, by the spatial colocalization of TFs and their target genes, may organize chromosomal domains, as suggested by a recent computational study [91]. To examine these possibilities, it would be desirable to analyze systematically the spatial distributions of TFs with high resolution in small bacteria cells. In this regard, recent advances in single-molecule based superresolution imaging methods such as PALM and STORM (Table 1.1) [36,37], offer great opportunities for bacteriologists [60,76,83].

Spatial distribution of RNA polymerase: are RNAP clusters active transcription centers?

Bacterial RNAP is a multi-subunit enzyme. The core enzyme consists of subunits $\alpha_2\beta\beta'\omega$, and is responsible for transcription elongation. The holoenzyme contains an additional σ factor, which is necessary for promoter recognition and initiation of transcription [92]. By using alternative σ factors, RNAP can be directed to transcribe different subsets of genes [93]. Currently all imaging studies probing the spatial distributions of RNAP in *E. coli* and *B. subtilis* have used fluorescent protein (FP) fusions of the largest subunit, β' (RpoC-FP). These fusion proteins have been shown to be able to replace the endogenous β', and >90% is incorporated into the RNAP core [94].

Using these fusion proteins, large and distinct RNAP foci have been observed in *E. coli* (Figure 1.5A) [52-54]. These foci were attributed to active transcription sites, reminiscent of eukaryotic transcription factories. Under superresolution imaging using RpoC-PAmCherry or RpoC-yGFP, these dense foci resolved into clusters ranging from 50 to 300 nm in diameter and often containing more than 100 localizations of RNAP [95,96]. As these clusters are largely diminished in cells growing in minimal media or treated with transcription inhibitors, they were

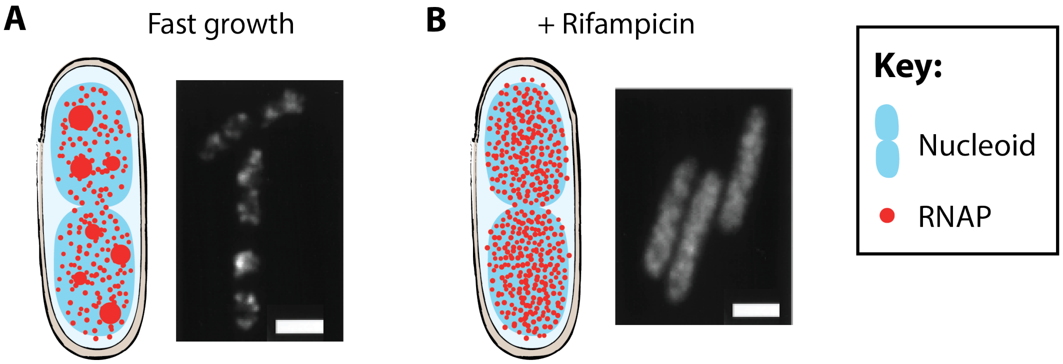


Figure 1.5 Spatial distribution of RNAP in E. coli. (A) Under fast growth, distinct foci of GFP-labeled RNAP can be seen. (B) Upon rifampicin treatment, RNAP distribution becomes more homogenous. Fluorescent images adapted from Cabrera et al. (2003), scale bars are 1 μm.

suggested to be active transcription centers (transcription factories) transcribing *rrn* operons (Figure 1.5B) [52-54,95]. Chromatin immunoprecipitation experiments supported this notion by showing that in rapidly growing cells RNAP predominately associates with ~ 90 transcription units that are involved in rRNA and ribosomal protein synthesis, and such associations are diminished in rifampicin treated cells [97,98]. Such a spatial arrangement, if proven true, would be reminiscent of the eukaryotic nucleolus, a specific sub-nuclear region where rRNA synthesis takes place.

One critical factor remaining unresolved for suggesting these foci as active transcription sites is that there is no direct evidence showing that RNAP molecules in observed foci are indeed actively engaged in transcription. Theoretically, these foci could be formed by RNAP molecules that are bound to DNA in nonproductive transcription complexes or on strong promoters poised for transcription [98]. In the latter cases, different growth conditions or drug treatment may alter the nucleoid structure, and hence alter the spatial distribution of RNAP bound on the nucleoid. Theoretical modeling and experimental quantifications have suggested that the majority of RNAP is associated with the nucleoid, but only ~ 20 - 30% of RNAP molecules are actively engaged in transcription, including both rRNA and mRNA synthesis [55,57,94,99,100]. Moreover, fluorescence recovery after photobleaching (FRAP) and single-molecule tracking experiments (Table 1.1) showed that in both rich and minimal media the majority of RNAP (50 - 80%) is stably associated with

DNA [95,101,102]. As the expression level and transcription activity of RNAP differ several fold under different growth conditions [55,57], these studies suggest that the decreased mobility of RNAP and the formation of foci may not necessarily be good indicators for transcription activity.

It is difficult to examine the spatial distribution of RNAP activity due to lack of specific drugs and antibodies targeting bacterial RNAP at different stages of transcription as that in eukaryotic cells. However, co-labeling RNAP with nascent RNAs using uridine analogs [103], transcription elongation factors such as NusG using antibodies [104], or genes using the superresolution FROS system as described above, may identify RNAP molecules that are engaged in active transcription. This will provide more concrete evidence to examine the hypothesis that these foci are active transcription factories.

The same critical analysis also applies to recent observations using fluorescence microscopy that RNAP mainly colocalizes with the nucleoid whereas ribosomes localize outside the nucleoid in *E. coli* and *B. subtilis* [54,96]. The latter is consistent with earlier EM evidence that ribosomes reside mostly outside of the nucleoid in *E. coli* [105]. However, the nucleoid-localized RNAP suggests that transcription and translation may be spatially separated, or decoupled, in these bacterial cells (note that in *C. crescentus* ribosome is distributed throughout the nucleoid but not at the periphery [106,107]). The spatial segregation between translation and transcription is in contrast to the long-held view supported by the classical Millar "Christmas Tree" spreads and many other biochemical studies

[108,109]. One way to reconcile these observations is that this large spatial separation between RNAP and ribosome does not necessarily represent a functional separation—as we have described above, not all RNAP molecules localized to the nucleoid are actively engaged in transcription. Because transcription of mRNA is only a small portion of total RNAP activity in fast growth conditions, it is possible that a small population of RNAP transcribing mRNAs is at the periphery of the nucleoid, where transcription and translation are coupled at the interface of RNAP and ribosome distributions. In addition, recently it was shown that the nucleoid has apparent voids in the middle of the cell, and hence it is possible that ribosome and RNAP are still coupled in these voids [110]. Alternatively, transcription-coupled translation could initiate inside the nucleoid, and when transcription is finished the mRNA detaches from the gene locus and moves outside of the nucleoid to finish translation.

Spatial distribution of RNAs: a combined result of translation and degradation

Where does an RNA molecule go when it is detached from its gene locus after transcription? Does it randomly diffuse inside the cell, localize to a specific cellular position, or just simply remain where it was transcribed? Moreover, is a particular localization pattern biologically important, or is it merely the result of physical constraints, without any biological consequences? To answer these questions, two important factors, translation and degradation of RNA, should be discussed. These two factors may significantly influence the spatial localization of RNAs.

Translation can influence the localization of mRNAs in two different ways. First, it is conceivable that with ongoing translation, the presence of multiple ribosomes and nascent peptide chains could significantly slow down the diffusion of mRNA purely due to the increased molecular size. Hence, translating mRNA molecules may exhibit limited mobility and largely remain where they are transcribed. Single-molecule tracking experiments using a fluorescently labeled ribosome subunit [111] or fluorescent protein fused bacteriophage MS2 coat protein MCP-ms2 labeling scheme (Table 1.1) [30,112] found that indeed 70S ribosomes and mRNAs displayed locally confined motion, although the diffusion constants (0.02 to 0.2 $\mu m^2/s$) are still large enough to enable random distributions over long time scales. In a few cases examined in *C. crescentus* and *E. coli*, both fluorescent in situ hybridization (FISH) (Table 1.1) on native mRNAs and the MCP-ms2 labeling (Table 1.1) showed that six different mRNAs encoding diverse protein products colocalized with their coding genes [107]. Such localization, as suggested by the authors, may be biologically significant—it could facilitate rapid interactions between proteins and their partners encoded in close gene clusters.

Note here that although these observations are consistent with the idea that translating ribosomes limit the mobility of mRNAs, other mechanisms could still be at play to localize (or delocalize) mRNAs. In *E. coli*, for example, a *lacZ* mRNA transcribed from a BAC plasmid carrying a disabled ribosome binding site still localized to cell poles and centers at which the plasmid resides [112]. In another study in *E. coli*, it was found that an untranslated RNA formed localized spots at cell poles, whereas a *lacZ* mRNA highly induced from the same multi-copy plasmid displayed largely homogenous distribution throughout the cells [113]. Whether these localization patterns are related to specific biological functions is unknown, but they demonstrate that the localization patterns may be dependent on the individual identity of RNA.

Another way for translation to influence the localization of mRNA is that co-translational insertion of nascent peptides into the membrane could target an mRNA to the membrane [114]. This has been demonstrated for three *E. coli* membrane proteins, lactose permease LacY, tetracycline efflux pump TetA and a glucose transporter $IICB^{Glc}$ [67,115] using fluorescence microscopy and membrane fractionation. In all three cases, the localization of mRNAs to the membrane is dependent on translation—frame-shifting or drug inhibition of translation abolished the localization. Interestingly, using the MCP-ms2 labeling scheme, another study reported similar membrane localization of the *E. coli* sugar permease BglF mRNA, but in a translation-independent way (Figure 1.6A) [116]. Most strikingly, the mRNAs coding for two additional proteins (a cytoplasmic enzyme BglB and a pole-localizing TF BglG) in the same polycistronic messenger, once decoupled from the *bglF* mRNA, exhibited drastically different localization patterns—*bglB* mRNA showed a helix-like pattern in the cytoplasm while *bglG* showed polar localization. The helix-like pattern was also observed for the cytoplasmic chloramphenicol acetyltransferase CAT mRNA [116]. Notably, these localization patterns were also detected in unlabeled mRNAs using FISH, arguing that they were not caused by the MCP-ms2 labeling scheme. The authors suggested that the localization of mRNAs to where their protein products are required could serve as a way to deliver proteins more efficiently to the desired locations; however, the translation-independent mechanism to transport these mRNAs to the cell pole or membrane remains unknown.

The degradation of RNA may also influence localization patterns. In *E. coli*, RNase E, the major enzyme that degrades RNAs [117], was reported to localize to the inner membrane or associate with cytoskeleton structures close to the membrane [118,119]. This spatial distribution suggests that RNAs could be targeted to membranes for degradation in *E. coli*. In fact, in the case described above for the *ptsG* mRNA coding for the glucose transporter IICB^{Glc}, it was found that membrane targeting enables efficient degradation of the mRNA with the help of a small RNA, which binds to the RNA chaperone, Hfq [115]. In *C. crescentus*, a small regulatory RNA called tmRNA was found to localize in a helix-like pattern in the cytoplasm when probed using FISH (Figure 1.6B) [120]. The tmRNA encodes the protein degradation tag ssrA in a process called trans-translation, in which the ssrA peptide is added on to a translating protein molecule stalled on the ribosome. Subsequently, the ssrA-tagged protein will be targeted for degradation and the stalled ribosome released [121]. Interestingly, the degradation enzyme for tmRNA in *C. crescentus*, RNase R, also forms a helix-like structure in the cytoplasm, but out of phase. The spatial segregation of RnaseR and tmRNA suggests that tmRNA may be protected from Rnase R degradation when its activity is needed [120].

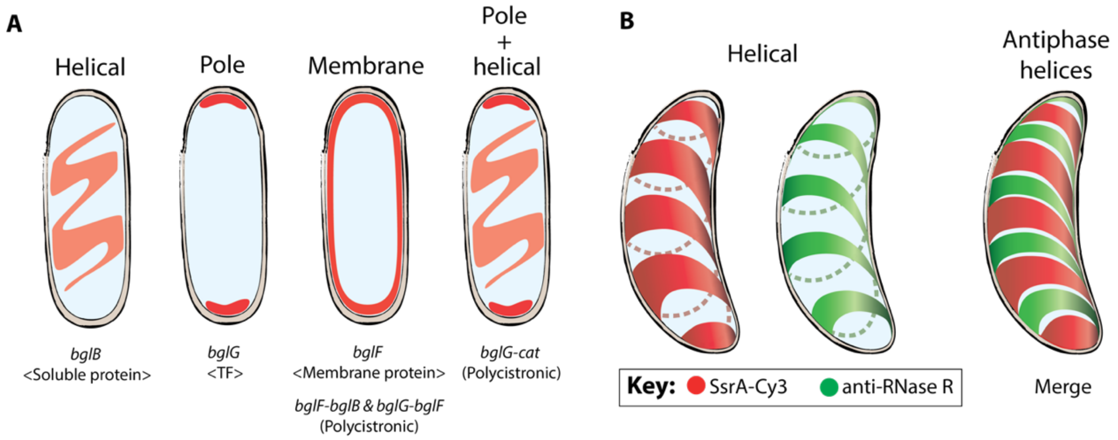


Figure 1.6 Distinct localization patterns of various bacterial mRNA. (A) Localization of mRNAs (labeled with the ms2-MCP system) coding for membrane protein BglF, cytoplasmic protein BglB and pole-localizing TF BglG show distinct cellular patterns. When in one polycistronic mRNA, *bglF* mRNA dominates the localization pattern of the cytoplasmic *bglB* and pole-localizing *bglF* to become membrane-localized (cell 3 from left). When the pole-localizing *bglG* mRNA and a cytoplasm-localizing *cat* mRNA are placed on the same polycistronic mRNA, the resulting localization pattern is mixed (cell 4 from left). (B) A small *C. crescentus* regulatory RNA, tmRNA forms a helix-like localization pattern in the cytoplasm (in red), and that of RNase R (which targets tmRNA for degradation) is localized in a similar anti-phase helical pattern (in green). tmRNA and RNase R are spatially segregated (merged).

Concluding remarks

As of today, the intracellular spatial distributions of genes, TFs, RNAP and RNAs have been documented with unprecedented sensitivity and resolution. However, knowing precisely where molecules are inside a cell is only the first step towards a better understanding of the connection between spatial organization and functionality. In analogy to Functional Magnetic Resonance Imaging, where brain activity can be visualized spatially, identifying the functional aspect of a particular spatial distribution pattern is critical. New imaging and biochemical methods focusing on correlating transcription activity with spatial localizations of various molecular species will no doubt shed light on the emerging question of the field—whether and how the spatial information is converted to regulatory signals of transcription.

This question is particularly important for bacterial cells, since they lack membrane compartmentalization for dedicated enzymatic tasks. Highly controlled but dynamic spatial organization of molecular components in a compact cell space may represent a new paradigm of regulation (Figure 1.7). In the canonical scheme for TF mediated gene regulation, the response is often mediated by signal-induced protein concentration changes through induction of protein expression at the transcription level, which occurs on a relatively long time scale. Reorganizing local environments by redistributing molecular components of transcription could be a potentially faster and more efficient regulatory mechanism to respond to a signal.

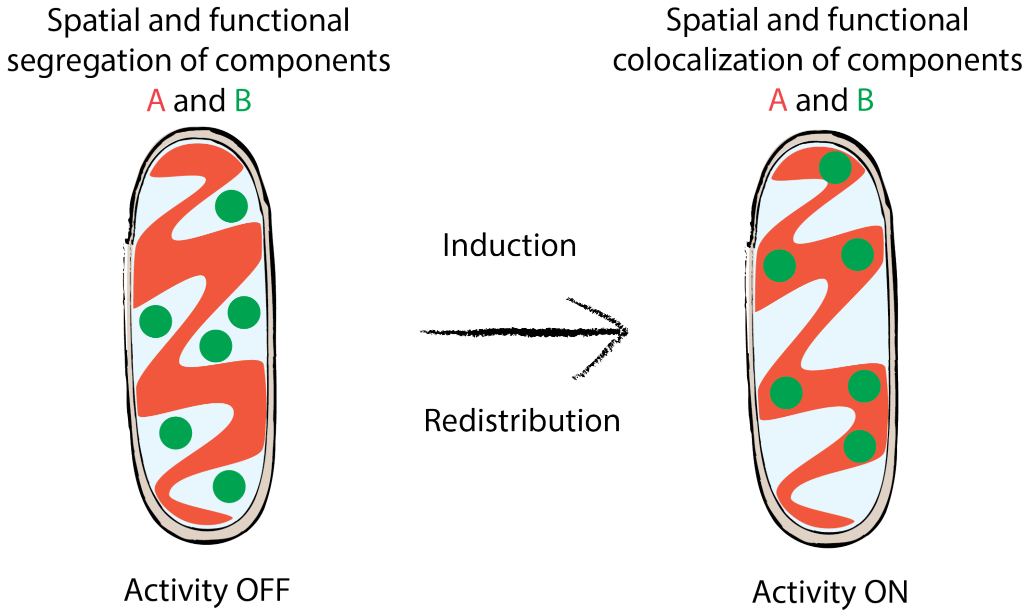


Figure 1.7 Model of transcription regulation by spatial organization. Molecular components of transcription can be spatially and functionally segregated. Upon induction, spatial reorganization of one or more of the components brings about a change in enzymatic activity.

Furthermore, spatial segregation of enzymes and substrates, such as the case for tmRNA degradation, could serve as a new way to regulate enzymatic activity without altering enzyme structure. New knowledge obtained from investigations into the spatial organization in bacteria will tie together previous known *in vitro* mechanisms of molecular interactions to produce a better understanding of this new type of regulation at the cellular level.

Chapter 2: The Spatial Organization of Transcription in *E. coli*

Introduction

Recent fluorescence-imaging studies have revealed that *E. coli* RNA polymerase (RNAP) is not homogenously distributed, but spatially clustered to form dense clusters during fast growth. Since rRNA synthesis is the major output of transcription under this condition [99], models have assigned the functional role of these RNAP clusters to the transcription of highly active genes such as the multiple ribosomal RNA operons (*rrn*) [52,53,96,122]. It has been proposed that multiple *rrn* operons spatially tether together with hundreds of RNAP to create transcription factories for efficient rRNA production, and that RNAP clusters are dependent on transcription activity [26]. Alternatively, models also propose that genes could be shuttled in and out of RNAP foci depending on their activity [123,124]. In eukaryotic cells, the classic example of membrane-free spatial and functional partitioning of enzymes is the clustering of RNAPI in the nucleolus for rRNA transcription. The transcription factory theory has also been more extensively investigated in eukaryotes for RNAPII, responsible for mRNA transcription. Nevertheless, whether RNAPII forms clusters is still controversial, with studies showing contradicting evidence for the presence of large functional clusters [47,51,125,126].

The transcription factory model in bacterial cells has not been directly verified, although a recent study in *E. coli* shows that six out of the seven *rrn* operons (all except the replication origin proximal *rrnC*) show high spatial colocalization of ~150-200 nm, independent of transcription activity from the *rrn* operons [127]. These experiments suggest that the 3D organization of the underlying chromosome structure exists independently of transcription activity. Naturally, these findings lead to the question of whether RNAP cluster organization is dependent on *rrn* operon transcription activity. Similarly, RNAP cluster formation could be independent of transcription activity and more closely tied to the global chromosome architecture which ultimately dictates the organization of RNAP binding sites.

In this study, we seek to investigate the functional role of RNAP spatial distribution and its dependence on high *rrn* transcription activity, and probe RNAP's spatial relationship with other components of transcription such as DNA sites, elongation factors and nascent transcripts in the bacterial model system *E. coli*. We found that while RNAP clusters were highly colocalized to rRNA synthesis sites during fast growth, both rRNA synthesis sites and RNAP clusters could exist independent of the other. We observed that the majority of RNAP clusters were retained under conditions where rRNA synthesis was diminished and in a mutant strain that lacked six out of the seven *rrn* operons [26]. Furthermore, we found that the RNAP clusters retained colocalization with *rrn* gene sites and elongation factor NusA with lowered rRNA transcription activity. These results indicate that a major component of RNAP cluster formation could be independent of rRNA transcription and may not be exclusive to *rrn* operons. As a follow-up, we investigated the colocalization of additional chromosomal sites with RNAP clusters and found that non-*rrn* segments could also colocalize to RNAP clusters. We additionally found that perturbations to the supercoiling state of the chromosome via gyrase inhibition shifted the cellular positioning of RNAP clusters while rRNA synthesis activity was retained. Moreover, with gyrase inhibition, the number of spatially distinct rRNA synthesis sites increased while the number of RNAP clusters stayed constant, further supporting the notion that rRNA transcription was not directly responsible for the formation of RNAP clusters. These findings indicate that chromosome structure is likely the major contributor on the formation and location of RNAP clusters inside the cell, and that the positioning/clustering of active rRNA operons is likely influenced by the supercoiling state of the chromosome.

Results

RNAP forms distinct clusters in cells growing in rich medium

To investigate the spatial organization of RNAP in *E. coli*, we used a strain in which the chromosomal *rpoC* gene encoding for the β’ subunit of RNAP is replaced by a photoactivatable fluorescent gene fusion, *rpoC-PAmCherry* [95,122,128]. We verified that the resulting RpoC-PAmCherry fusion protein was expressed as full-length (Figure 2.1A), was incorporated efficiently into the RNAP core enzyme complex (Figure 2.1B), and supported normal cell growth as the sole cellular source of β’ subunit (Figure 2.1C). Therefore, the spatial distribution and dynamics of the RpoC-PAmCherry fusion protein should be representative of the RNAP core or holoenzyme. In the text below for simplicity, we refer to this fusion strain as RNAP-PAmCherry.

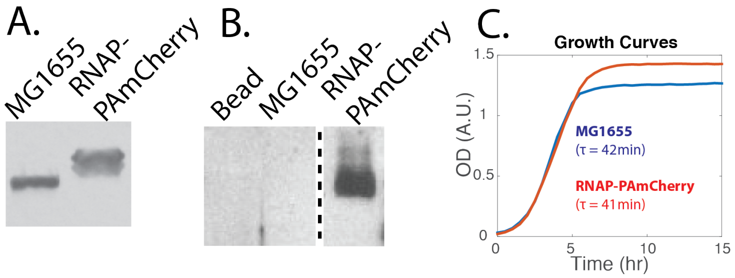


Figure 2.1 RpoC-PAmCherry was expressed in full-length and supports normal cell growth. (A) RpoC-PAmCherry showed correct molecular size as a full-length fusion (detected using α-RpoC in a western blot), the MG1655 strain is the wild-type strain. (B) Co-immunoprecipitation of RNAP core enzyme from *E. coli* cell extract of wild-type MG1655 and RNAP-PAmCherry cells using bead-conjugated α-RpoB and detected using α-mCherry. (C) Growth curves (EZRDM, 30C) showed no significant difference in cell doubling times between MG1655 and RNAP-PAmCherry strains.

Using RNAP-PAmCherry, we first performed single-molecule localization based superresolution imaging [36] (Figure 2.2) on exponentially growing cells in EZ Rich Defined Media (EZRDM) at room temperature (RT). We observed clustered distributions of RNAP-PAmCherry (Figure 2.3A) in individual cells that were similar to what has been reported previously under similar growth conditions [95,96]. To quantitatively characterize RNAP clusters, we performed a density-based threshold analysis to isolate RNAP clusters (Figure 2.3B, see materials and methods for details). We verified that the clustered distribution of RNAP-PAmCherry was not due to the weak dimerization property of PAmCherry, because RpoC fused with a true monomeric mEos3.2 [129], or Immunofluorescence superresolution imaging [37] using an antibody against the endogenous β subunit (RpoB) or β’ subunit (RpoC) in wildtype MG1655 cells, all showed similar clustered distributions of RNAP (Figure 2.3C, D, and E). Live-cell PALM images of RNAP-PAmCherry and RNAP-mEos3.2 showed an identical number of RNAP clusters per cell (see cluster characterization below), and these clusters are not a result of a 2D projection of a random distribution in the 3D cell volume (Figure 2.3F, G). The clustered distribution was also not due to the blinking behavior of PAmCherry, as we have developed a robust computational method to remove multiple localizations of the same PAmCherry molecules (Bohrer et al., 2018, manuscript in preparation). Finally, we verified that RNAP-PAmCherry clusters were bound to chromosomal DNA instead of being aberrant protein aggregates, as overexpression of an anti-σ factor AsiA [130] resulted in the diminishing of RNAP clusters, indicating that σ^{70} was needed for optimal RNAP binding and cluster formation (Figure 2.3H, F). Collectively, these data indicate that the clustered distribution we observe for RNAP-PAmCherry reflect the underlying spatial distribution of RNAP in live *E. coli* cells.

We found that under the rich medium growth condition (EZRDM, cell doubling time = 41 ± 2 min, Figure 2.1C), on average each cell had ~2 dense RNAP clusters (2.13 ± 0.05, n= 664, Figure 2.4A, Table 2.1), containing ~ 16% of total RNAP-PAmCherry molecules detected (0.16 ± 0.05, n = 664, Figure 2.4B, Table 2.1), with each cluster containing ~ 8% of total RNAP molecules (0.076 ± 0.001, n = 1385, Figure 2.4C, Table 2.1); this result is significantly different from what would be expected from a random distribution of RNAP (Figure 2.4A, B and C, black curves, Figure 2.3G, Table 2.1).

Previous studies have shown that treating cells with rifampicin (Rif) leads to the dispersion of RNAP clusters [53]. Rifampicin is a global transcription inhibitor and prevents initiating RNAP molecules from escaping promoters, but does not affect elongating RNAP [25,131]. Indeed, in cells treated with rifampicin we observed both a significant downshift of the number of RNAP clusters and the fraction of cluster-localized RNAP molecules per cell (Figure 2.4 A, B, and C). However, some RNAP clusters persisted (Figure 2.4D), and the distribution pattern was still significantly different from what would be expected from a random distribution of RNAP (Figure 2.4A, B and C, black curves).

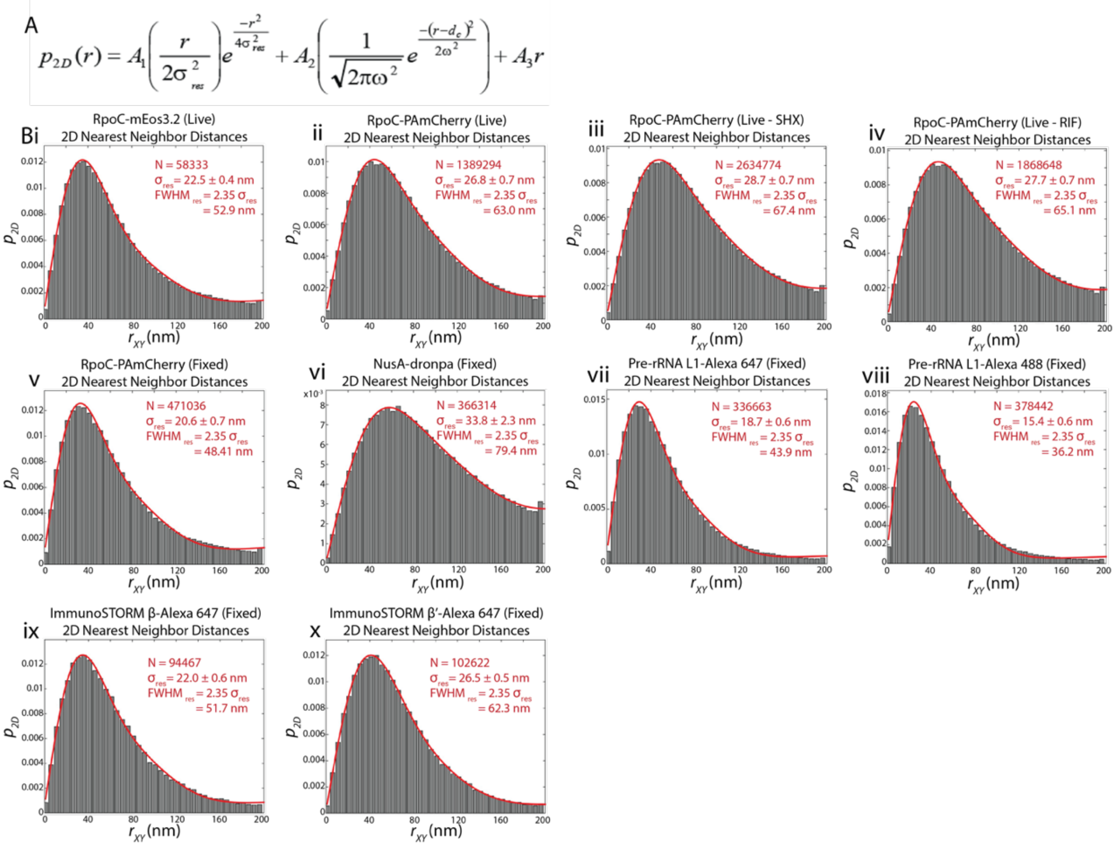


Figure 2.2 Measurement of spatial resolution. (A) Equation describing the distribution (p) of distances (r) between the nearest neighbors in adjacent frames of PALM data with the localization precision σ_{res}, using two dimensions (p_{2D}). This equation was adapted by Endesfelder *et al.* from the 2D distance distribution expected of repeat localizations from the same molecule to account for the possibility that one molecule's nearest neighbor in the adjacent frame may be another molecule. The adaption is in the 2nd and 3rd terms of the equation in (A) by the Gaussian parameters ω and d_c, and the weight factors A_1, A_2, and A_3. (B) XY distances between nearest neighbor localizations in adjacent frames were calculated for a subset of data representing all listed experimental conditions. (Bi-x) The gray bars show the distribution of r_{XY}, the number of distances used is listed as N. The fit (red) with equation (A) yield a Gaussian localization precision σ_{res}, which begets a spatial resolution $FWHM_{res}$ (FWHM of the Gaussian localization precision), both are listed accordingly.

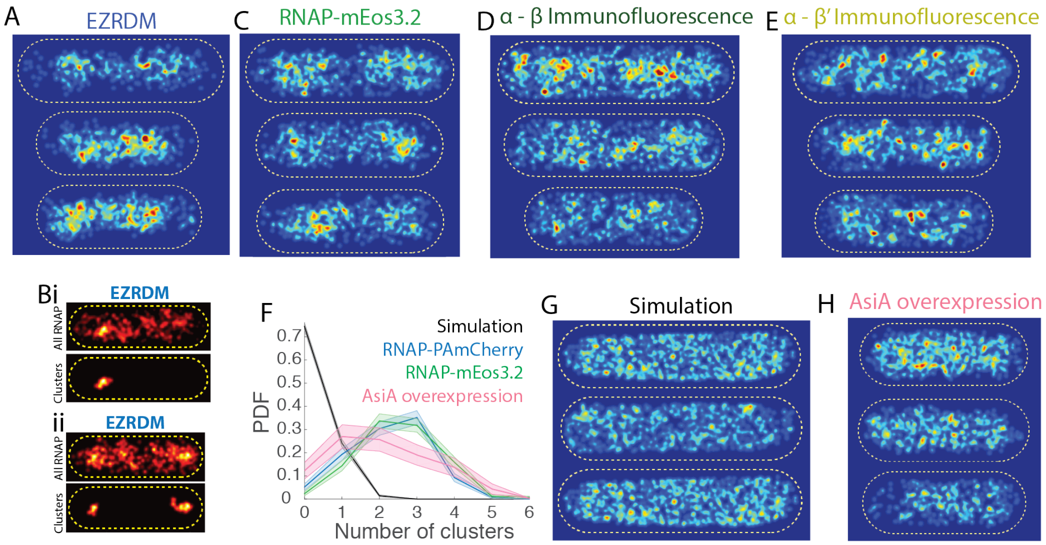


Figure 2.3 RNAP clusters were not a result of FP fusion oligomerization. (A) Example reconstructed superresolution images of RNAP (RpoC-PAmCherry) under rich medium growth (EZRDM). (B) Cluster identification using modified tree-clustering analysis to threshold for high density regions, representative cells from EZRDM condition are shown, where (i) is a cell with a single cluster, and (ii) is a cell with two clusters. (C) Representative reconstructed superresolution images of RNAP (RpoC-mEos3.2) under EZRDM. Representative reconstructed superresolution images of RNAP in wild-type cells probed via α-β (RpoB) (D), and α-β' (RpoC) (E) under EZRDM. (F) Number of clusters per cell distribution comparison between live cell experiments: RpoC-PAmCherry, RpoC-mEos3.2 and RpoC-PAmCherry/AsiA overexpression, with negative simulations shown in black, all errors are standard errors from bootstrapping. (G) Example of negative simulation of randomly distributed RNAP in the cell volume, see materials and methods for details on simulation. (H) Representative reconstructed superresolution images of RNAP (RpoC-PAmCherry) under EZRDM growth with AsiA (σ^{70}) overexpression.

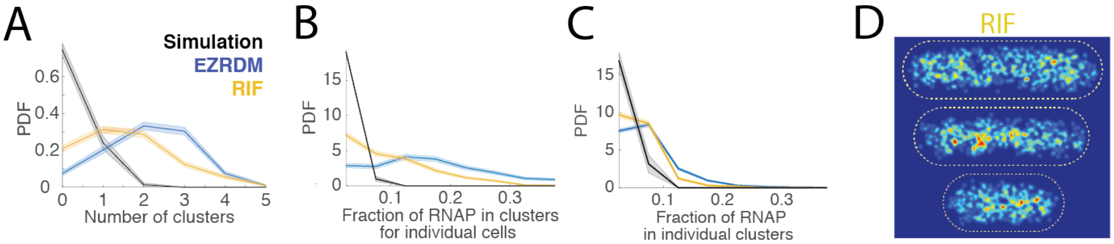


Figure 2.4 Characterization of RNAP clusters in live cells. (A) Number of RNAP clusters per cell distribution. (B) Fraction of localizations within clusters per cell distribution. (C) Fraction of localizations within clusters distribution. For A-C, black curve shows results of cluster analysis on simulated random distribution within the same experimental cell volume (3D simulation projected into 2D), all errors are standard errors from bootstrapping. (D) Representative reconstructed superresolution images of RNAP in rifampicin treated cells.

Table 2.1 RNAP cluster characteristics in live cell PALM images for various conditions. *Values shown are mean ± standard error, with (N) as the number of data points used.*

Condition (live cell)	**# of RNAP clusters per cell**	**Fraction in clusters per cluster**	**Fraction in clusters per cell**
Neg. simulation	0.27 ± 0.04 (664)	0.044 ± 0.001 (1385)	0.01 ± 0.002 (664)
EZRDM	2.13 ± 0.05 (664)	0.076 ± 0.001 (1385)	0.16 ± 0.005 (664)
SHX	1.85 ± 0.04 (714)	0.075 ± 0.001 (1317)	0.14 ± 0.004 (714)
RIF	1.54 ± 0.05 (559)	0.061 ± 0.001 (860)	0.09 ± 0.003 (559)
Δ6*rrn*	1.83 ± 0.08 (151)	0.070 ± 0.002 (277)	0.13 ± 0.006 (151)
EZRDM (mEos3.2)	2.51 ± 0.07 (258)	0.086 ± 0.002 (648)	0.22 ± 0.007 (258)
AsiA overexpression	2.07 ± 0.16 (75)	0.065 ± 0.003 (155)	0.13 ± 0.011 (75)

RNAP clusters colocalized with the *rrnD* operon, nascent rRNA, and NusA in cells growing in rich medium

To directly probe whether the clustered organization of RNAP is dependent on *rrn* transcription activity, we investigated whether RNAP clusters are transcription centers actively engaged in rRNA synthesis. We first examined whether RNAP clusters colocalized with an *rrn* operon. We used a high-resolution chromosomal DNA marker previously developed in the lab (Figure 2.5A) to label the *rrnD* operon, and performed two-color superresolution imaging simultaneously with RNAP-PAmCherry (Figure 2.5B). We found that ~50% of *rrnD* (0.52 ± 0.07, n=186, Figure 2.5C, Table 2.2) colocalized with RNAP clusters within a 100-nm radius threshold (the experimental localization precision is ~30nm, see Figure 2.2), significantly higher than what is expected from a simulated random colocalization pattern (Figure 2.5C, thin black bars). For details on colocalization calculation, see materials and methods.

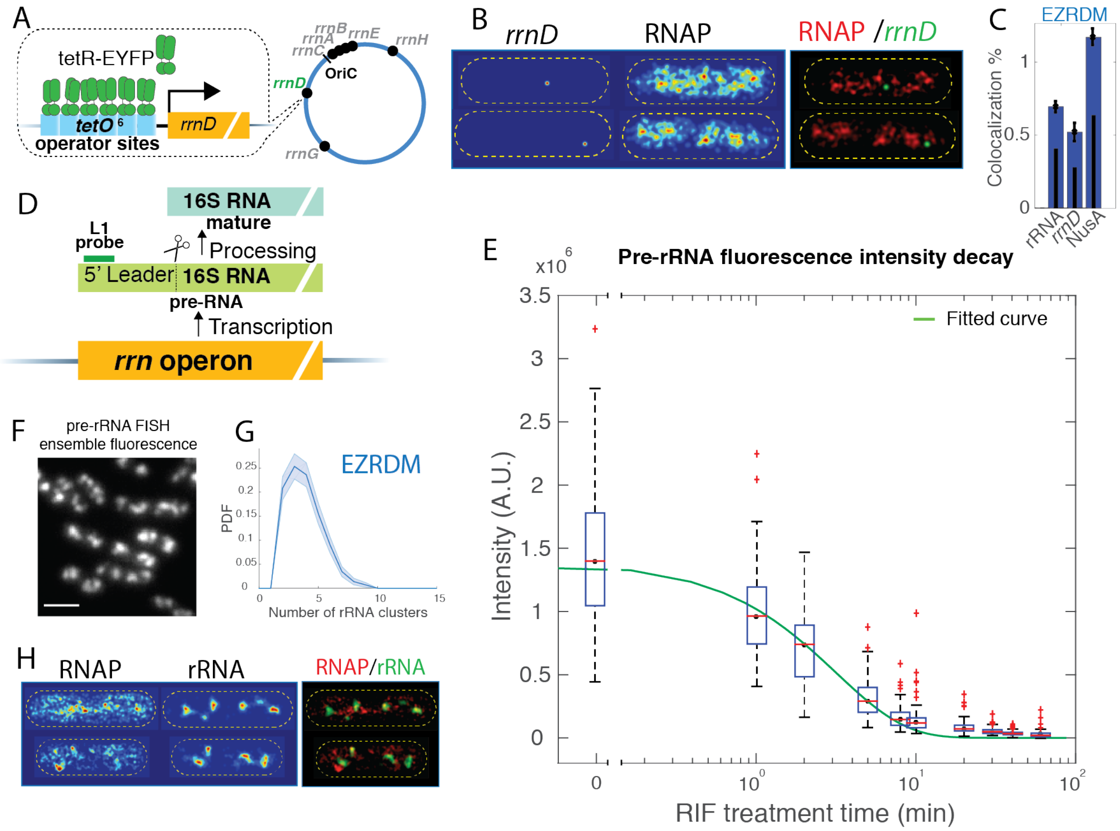


Figure 2.5 Colocalization of RNAP clusters with other transcription components under EZRDM condition. (A) Schematic for DNA labeling and *rrn* operons, bubble shows modified FROS system for chromosome site labeling. (B) Representative two-color superresolution images of *rrnD* and RNAP, with overlay shown on the right. (C) Colocalization values (100-nm distance threshold) for *rrnD*, (50-nm distance threshold) rRNA clusters and (50-nm threshold) NusA clusters to RNAP clusters, errors are standard errors, negative simulation results are shown as thin black bars, for details on colocalization analysis see materials and methods). (D) Pre-rRNA detection scheme, L1 probe was used to label nascent pre-rRNA prior to their processing and degradation. (E) Ensemble fluoresence of pre-rRNA FISH signal for single cells are measured at each time point of RIF treatment, which stops transcription globally. The distribution of fluorescence is plotted as bar plots, with the population mean in red, and boxed region as the 25th and 75th percentiles. The fluorescence decay was fit with a single exponential with a decay constant of 0.32, giving a half-life of ~130 sec. (0min (n = 166), 1min (n = 76), 2min (n = 73), 5min (n = 88), 8min (n = 84), 10min (n = 83), 20min (n = 89), 30min (n = 110), 40min (n = 103), 60min (n = 96)). (F) Pre-rRNA ensemble fluorescence (large field view) in cell under EZRDM condition, scale bar is 2μm. (G) Distribution of the number of rRNA clusters per cell for the EZRDM condition. (H) Representative superresolution images of RNAP and rRNA in cells grown in EZRDM, with overlay shown on the right.

Table 2.2 DNA site colocalization values with RNAP clusters in live cell PALM images for various conditions. *Values shown are mean ± standard error. All colocalization values used a 100-nm distance threshold, with (N) as the number of data points used in measurements, all simulations use the same N as the corresponding experimental data.*

DNA site	EZRDM	EZRDM (Sim)	SHX	SHX (Sim)	RIF	RIF (Sim)	Galactose	Galactose (Sim)
rrnD	0.52 ± 0.07 (186)	0.28 ± 0.06	0.33 ± 0.03 (211)	0.22 ± 0.05	0.24 ± 0.06 (140)	0.22 ± 0.06	-	-
rrnC	0.39 ± 0.05 (145)	0.26 ± 0.06	0.35 ± 0.05 (138)	0.22 ± 0.05	0.35 ± 0.06 (136)	0.24 ± 0.06	-	-
rrnH	0.35 ± 0.05 (258)	0.26 ± 0.05	0.32 ± 0.04 (333)	0.28 ± 0.04	0.29 ± 0.06 (207)	0.23 ± 0.04	-	-
rrnG	0.25 ± 0.04 (152)	0.22 ± 0.06	0.34 ± 0.04 (313)	0.28 ± 0.05	0.28 ± 0.04 (200)	0.24 ± 0.04	-	-
galE	0.50 ± 0.06 (131)	0.25 ± 0.06	0.35 ± 0.03 (341)	0.25 ± 0.03	0.46 ± 0.07 (107)	0.27 ± 0.05	0.39 ± 0.06 (165)	0.26 ± 0.05
mraZ	0.38 ± 0.04 (221)	0.25 ± 0.05	-	-	-	-	-	-
yegH	0.31 ± 0.04 (128)	0.23 ± 0.05	-	-	-	-	-	-

Next, we directly probed the colocalization between RNAP clusters with nascent rRNA synthesis sites, as colocalization with the *rrn* operon does not necessarily indicate that these RNAP clusters are active in transcription. We labeled nascent pre-rRNA using a dye-conjugated oligonucleotide targeting the 5' leader region of the 16S pre-rRNA, which is absent from mature 16s RNA inside ribosomes [132] (Figure 2.5D). The 5' leader degrades rapidly after being processed, with a half-life of ~130 sec (Figure 2.5E). The pre-rRNA signals displayed clear spot-like puncta with an average of ~4 foci per cell (3.86 ± 0.09, n = 288, Figure 2.5F, Figure 2.5G, Table 2.3), suggesting that rRNA synthesis sites are spatially confined to approximately two sites per nucleoid, consistent with a recent study demonstrating the spatial colocalization of most *rrn* operons [127]. Importantly, two-color superresolution imaging of pre-rRNA with RNAP-PAmCherry (Figure 2.5H) showed that ~80% of RNAP clusters (0.83 ± 0.02, n = 404, Table 2.4) colocalized with a nascent rRNA cluster signal within a 50-nm radius. By comparison, ~ 70% of rRNA clusters (0.69 ± 0.07, n = 720, Figure 2.5C, Table 2.4) colocalized with

RNAP clusters within a 50-nm distance threshold. Thus, most RNAP clusters were likely engaged in active rRNA synthesis, and the vast majority of rRNA synthesis sites resided within or are proximal to RNAP clusters.

Table 2.3: rRNA cluster and RNAP cluster characteristics in rRNA-RNAP two-color experiments (fixed cell). *Values shown are mean ± standard error, with (N). There was a lack of signal for SHX condition for rRNA, due to low rRNA synthesis activity, some rRNA cluster parameters were not calculated.*

Condition	# of RNAP clusters	# of rRNA clusters	Fraction in cluster per cluster (rRNA)	Fraction in clusters per cell (rRNA)
Neg. Simulation	0.06 ± 0.02			
EZRDM	2.07 ± 0.06 (400)	3.86 ± 0.09 (288)	0.16 ± 0.004 (1086)	0.63 ± 0.005 (288)
SHX	0.96 ± 0.09 (142)	0.12 ± 0.02 (12)	-	-
Δ6*rrn*	1.96 ± 0.07 (285)	4.95 ± 0.12 (192)	0.14 ± 0.003 (939)	0.70 ± 0.006 (192)
Nalidixic acid	2.10 ± 0.12 (108)	6.02 ± 0.22 (99)	0.09 ± 0.003 (583)	0.53 ± 0.008 (99)
Novobiocin	1.71 ± 0.11 (86)	5.53 ± 0.17 (153)	0.09 ± 0.003 (853)	0.52 ± 0.006 (153)

Table 2.4: rRNA cluster colocalization values with RNAP clusters in fixed cell PALM images for various conditions. *Values shown are mean ± standard error. All colocalization values used a 50-nm distance threshold, with (N), all simulations use the same N as the corresponding experimental data.*

Condition	rRNA to RNAP	rRNA to RNAP (Sim)	RNAP to rRNA	RNAP to rRNA (Sim)
EZRDM	0.69 ± 0.03 (720)	0.32 ± 0.03	0.83 ± 0.02 (404)	0.42 ± 0.02
Δ6*rrn*	0.68 ± 0.04 (586)	0.35 ± 0.03	0.76 ± 0.03 (247)	0.40 ± 0.03
Nalidixic acid	0.60 ± 0.05 (476)	0.39 ± 0.04	0.77 ± 0.04 (183)	0.46 ± 0.04
Novobiocin	0.52 ± 0.05 (378)	0.33 ± 0.04	0.79 ± 0.04 (148)	0.49 ± 0.05

To further probe what other molecular components are in RNAP clusters, we examined the colocalization of RNAP clusters with the essential elongation factor NusA. NusA is an anti-termination factor for *rrn* operons; it binds RNAP quickly after transcription initiation [133] and is essential for the transcription of full length pre-rRNAs [134-136]. On other operons, NusA can also stimulate pausing or transcription termination [133,137-139]. We generated a dual-labeled strain in which the endogenous *nusA* gene was replaced by a *nusA-dronpa* fusion gene in the background of the *rpoC-PAmCherry* fusion (Figure 2.6A). We verified that the NusA-Dronpa fusion protein fully complemented *nusA* null phenotype, and the dual-labeling strain grew at a similar growth rate compared to WT cells (Figure 2.6B); for simplicity, we will refer to this strain as RNAP-PAmCherry/NusA-Dronpa. Two-color superresolution imaging of RNAP-PAmCherry/NusA-Dronpa showed that NusA formed a similar clustered distribution (Figure 2.6C), and colocalized completely with RNAP clusters within a 50-nm radius (1.19 ± 0.04, n = 207, Figure 2.5C, Table 2.5). Note here that the large size of NusA clusters resulted in a larger than one colocalization value. Similarly, RNAP clusters also had complete colocalization with rRNA clusters within a 50-nm radius (0.98 ± 0.02, n = 232, Table 2.5).

Taken together, our results demonstrated that in cells growing in rich medium, there were high levels of colocalization of RNAP clusters with the *rrnD* operon, nascent rRNA synthesis sites, and the transcription factor NusA, supporting the transcription factory model under this growth condition.

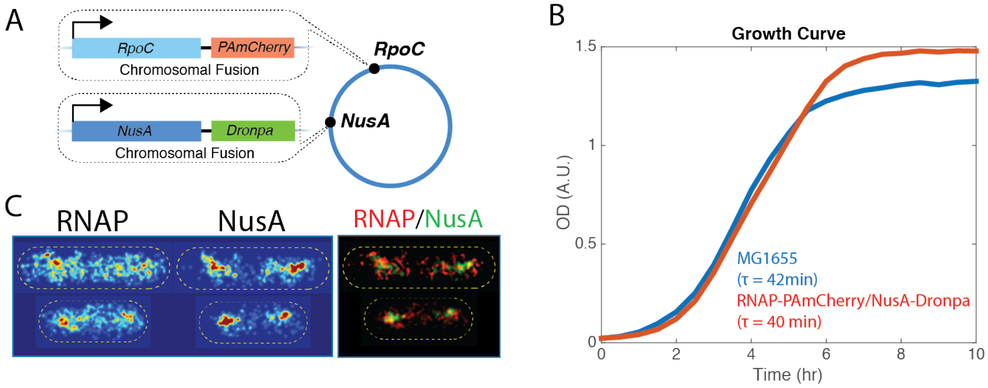


Figure 2.6 NusA-Dronpa exhibited a clustered distribution. (A) NusA-Dronpa and RpoC-PAmCherry labeling scheme on the chromosome. (B) Growth curves (EZRDM 30C) showed no significant difference in cell doubling times between wildtype MG1655 and the dual fusion RNAP-PAmCherry/NusA-Dronpa strains. (C) Representative superresolution images of RNAP and NusA in cells grown in EZRDM, with overlay shown on the right.

Table 2.5 NusA cluster colocalization values with RNAP clusters in fixed cell PALM images for various conditions. *All colocalization values used a 50-nm distance threshold, all simulations use the same (N) as the corresponding experimental data.*

Condition	**NusA to RNAP**	**NusA to RNAP (Sim)**	**RNAP to NusA**	**RNAP to NusA (Sim)**
EZRDM	1.19 ± 0.04 (207)	0.57 ± 0.05	0.98 ± 0.02 (232)	0.48 ± 0.04
SHX	1.22 ± 0.03 (289)	0.71 ± 0.04	0.95 ± 0.01 (347)	0.51 ± 0.03

Colocalization of RNAP clusters with the *rrnD* operon and NusA was not dependent on active rRNA synthesis

Our next question asked whether the spatial clustering of RNAP was dependent on active rRNA transcription as proposed in the transcription factory model. As a first step, we treated cells with serine hydroxamate (SHX), a seryl-tRNA synthetase inhibitor [134,140,141]. SHX induces the stringent response, during which amino acid starvation leads to the production of ppGpp, a small molecule messenger that binds to RNAP directly to destabilize open complex formation on the P1 promoters of *rrn* operons, thereby effectively inhibiting rRNA synthesis [140,141]. Consistent with this role, we observed a dramatic reduction in total rRNA synthesis in SHX treated cells via pre-rRNA FISH, only ~3% of residual rRNA synthesis remained, likely from *rrn* P2 promoters as previously reported [142] (Figure 2.7A, B). In contrast to the significant reduction of rRNA synthesis, we observed that majority of RNAP clusters remained (Figure 2.7C), with only a slight but significant reduction in the number of RNAP clusters, fraction of RNAP in individual clusters, and the fraction of RNAP molecules in clusters per cell (Figure 2.7Di-iii, Table 2.1). Interestingly, two-color superresolution imaging of RNAP clusters in SHX-treated cells with *rrnD* operon and NusA showed that the colocalization levels with *rrnD* operon decreased but were still higher than randomly simulated control (0.33 ± 0.03, N = 211, Figure 2.7E, F, Table 2.2). Moreover, colocalization between NusA clusters and RNAP clusters did not decrease (1.22 ± 0.03, N = 289, Figure 2.7E, G, Table 2.5). These results suggested that under this condition, some RNAP clusters were not engaged in active rRNA synthesis, but are still colocalized with the *rrnD* operon and the transcription factor NusA.

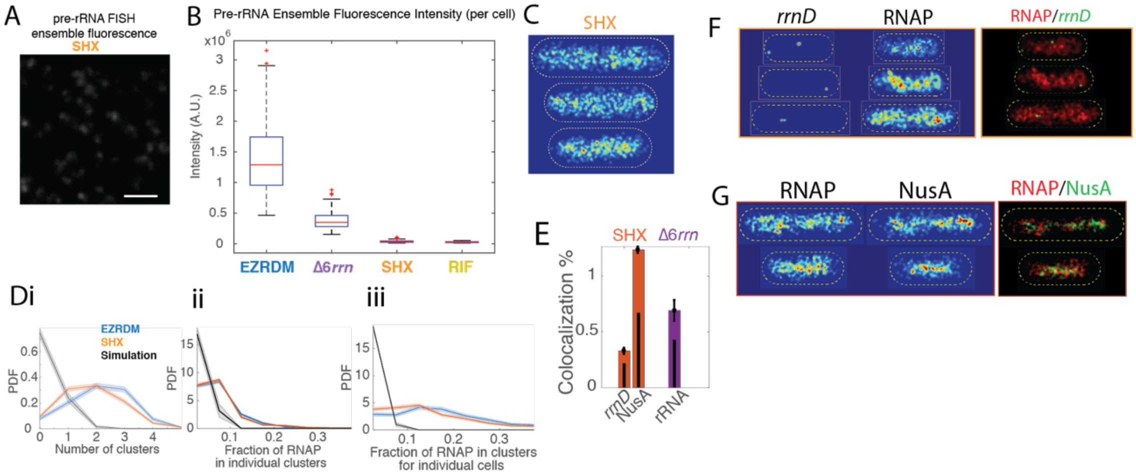


Figure 2.7 Colocalization of RNAP clusters with other transcription components in SHX treated cells. (A) Pre-rRNA ensemble fluorescence (large field view) for cells treated with SHX, scale bar is 2μm. (B) Ensemble fluoresence of pre-rRNA FISH signal for single cells are measured for each condition after cell segmentation. The distribution of cellular FISH signal in the different conditions are plotted as bar plots, with the mean in red, and boxed region as the 25th and 75th percentiles, (EZRDM (n = 72), Δ6*rrn* (n = 72), SHX (n = 110), RIF (n = 76)). (C) Representative superresolution images of RNAP in cells treated with SHX. (D) RNAP cluster characterizations for SHX treated cells: (i) number of clusters per cell, (ii) fraction of RNAP in individual clusters and (iii) Fraction of RNAP in clusters for individual cells. (E) Colocalization values between *rrnD* (100-nm threshold) and NusA clusters (50-nm threshold) with RNAP clusters in cells treated with SHX; colocalization between rRNA clusters and RNAP clusters in Δ6*rrn* cells (50-nm threshold), errors are standard errors, negative simulation results are shown as thin black bars. (F) Representative two-color images of *rrnD* and RNAP in SHX treated cells. (G) Representative two-color images of RNAP and NusA in SHX treated cells.

We reasoned that because SHX treatment could lead to global changes in cell physiology and nucleoid structure, the results we observed on the changes of RNAP clusters might not be directly due to changes in rRNA synthesis activity. Therefore, to probe more directly the influence of rRNA synthesis on RNAP clusters, we used a strain, Δ6*rrn*, in which six out of seven *rrn* operons (except for *rrnC*) are deleted [26]. The Δ6*rrn* strain grew with slightly slower rate compared to wildtype cells under the same rich medium growth condition (cell doubling time = 47 ± 2 min, Figure 2.8A), and showed a significant reduction in total rRNA synthesis via pre-rRNA FISH (Figure 2.8B, Figure 2.7B). However, similar to SHX treated cells, RNAP clusters largely persisted (Figure 2.8C, Di-iii). While rRNA clusters in Δ6*rrn* colocalized with RNAP clusters to a similar degree compared to wildtype RNAP-PAmCherry (~70%), there was a ~7% decrease in the colocalization of RNAP clusters to rRNA clusters, indicating that a portion of RNAP clusters in Δ6*rrn* may not be engaged in synthesis of rRNA (Figure 2.7E, Figure 2.8E, Table 2.4). These results further suggested that while the majority of RNAP clusters under fast growth are colocalized with actively engaged rRNA synthesis, RNAP's clustered organization may not be dependent on the transcription activity of *rrn*.

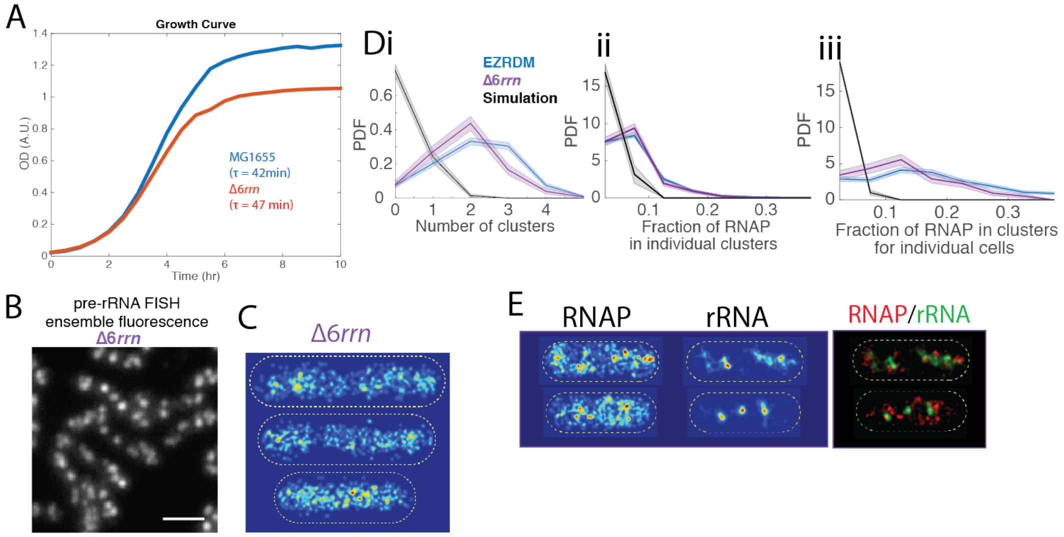


Figure 2.8 Characterization of RNAP clusters in Δ6*rrn* cells. (A) Growth curve (EZRDM 30C) of Δ6*rrn* strain compared to wildtype MG1655. (B) Pre-rRNA ensemble fluorescence (large field view) for Δ6*rrn* cells, scale bar is 2μm. (C) Representative superresolution images of RNAP in Δ6*rrn* cells. (D) RNAP cluster characterizations for Δ6*rrn* cells: (i) number of clusters per cell, (ii) fraction of RNAP in individual clusters and (iii) Fraction of RNAP in clusters for individual cells. (E) Representative two-color images of RNAP and rRNA in Δ6*rrn* cells.

Changing the supercoiling state of the chromosome via global inhibition of gyrase causes redistribution of RNAP clusters and rRNA synthesis sites

Finally, to investigate whether global chromosome architecture is important for RNAP distribution, in particular whether RNAP clusters are perturbed by changes in DNA supercoiling, we performed drug inhibition of gyrase while looking at the distribution of RNAP and nascent pre-rRNA simultaneously. In *E. coli*, gyrase is responsible for removing positive supercoils ahead of active replication and transcription; the inhibition of gyrase has been shown to affect the transcription of many genes and cause an overall down-shift in RNA production [143-148]. To avoid long-term detrimental effects such as dsDNA breakage, which can be caused by gyrase inhibition, we treated cells with two types of gyrase inhibitors - either nalidixic acid (Nal) or novobiocin [149,150] - for a short period of time (10-30 min). Interestingly, there was no major perturbation to nascent rRNA production based on the total cellular intensity measured from ensemble fluorescence images (Figure 2.9A). We determined that this result is not likely due to changes in rRNA degradation, as we saw that neither Nal nor novobiocin caused significant changes in the pre-rRNA degradation rate (Figure 2.9B). We also performed experiments with gyrase inhibition for up to 150 min and with higher novobiocin drug concentrations, observing fairly steady rRNA production in all cases (Figure 2.9C).

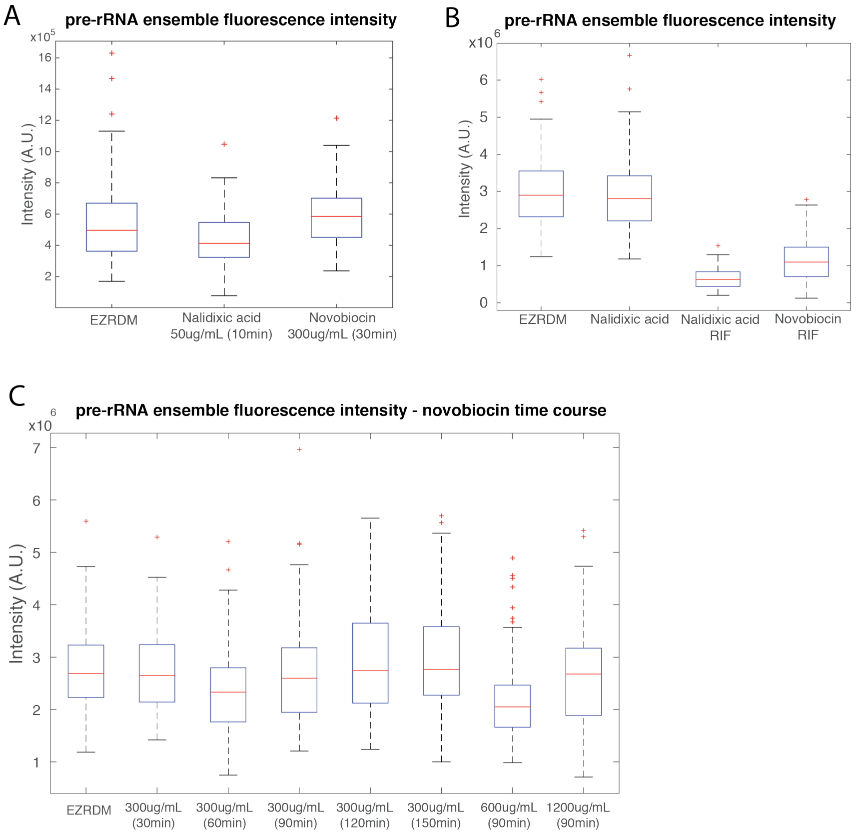


Figure 2.9 Pre-rRNA ensemble fluorescence intensities under gyrase inhibited conditions. (A) Quantification of cellular pre-rRNA ensemble fluoresence signal intensities for EZRDM, nalidixic acid (10 min 50ug/mL) and novobiocin (30min 300ug/mL) conditions. (B) Quantification of cellular pre-rRNA signal intensities for EZRDM, nalidixic acid (10 min 50ug/mL), and two gyrase treatment conditions with a 10-min rifampicin treatment follow-up (100ug/mL) without washing out the gyrase inhibitors. (C) A time series of novobiocin treatment (0 – 150 min, 300ug/mL), with higher concentration of novobiocin also used (600ug/mL and 1200ug/mL, 90 min). Ensemble fluorescence intensities were collected from ~100 cells for each condition presented. (A-C) Mean is shown in red, with 25th and 75th percentiles are shown by the boxed region.

Both NaI and novobiocin treated cells showed a clear difference in their pre-rRNA cluster characteristics in comparison to the fast growth condition. There were ~40-50% more pre-rRNA clusters per cell in gyrase-inhibited cells compared to the fast growth condition (NaI: 6.02 ± 0.22, n = 99, novobiocin: 5.53 ± 0.17, n = 153, Figure 2.10A, Table 2.3). These pre-rRNA clusters had on average about half of the fraction of pre-rRNA in individual clusters, and a ~20% decrease in fraction of rRNA in clusters per cell compared to the fast growth condition (Table 2.3, Figure 2.10B, C). These results indicated that gyrase treated cells had a greater number of small more dispersed pre-rRNA clusters (Figure 2.11A, B, C, D). Strikingly, these pre-rRNA clusters were located closer to the cell periphery on the short axis for both Nal and novobiocin treatments compared to fast growing wildtype cells and Δ6*rrn* (Figure, 2.11D, E). The gyrase inhibition experiments were the only cases where we saw a clear shift in cellular positioning of the pre-rRNA clusters. Similar to what we have observed in SHX-treated or the *Δ6rrn* cells, we observed that the RNAP clusters were retained under gyrase inhibition (Nal: 2.1 ± 0.12, n = 108, novobiocin: 1.71 ± 0.11, n = 86, Table 2.3, Figure 2.11F). In correlation, RNAP clusters under gyrase treatment shifted towards the cell periphery on the short axis compared to the previously tested conditions of fast growth, SHX and *Δ6rrn* (Figure 2.11B, C, G, H).

For the gyrase inhibited conditions, the pre-rRNA clusters and RNAP clusters had a slightly decreased level of colocalization compared to the fast growth condition. This behavior is likely due to the spatial dispersion and the increased number of pre-rRNA clusters in the gyrase inhibited condition (Figure 2.12, Table 2.4).

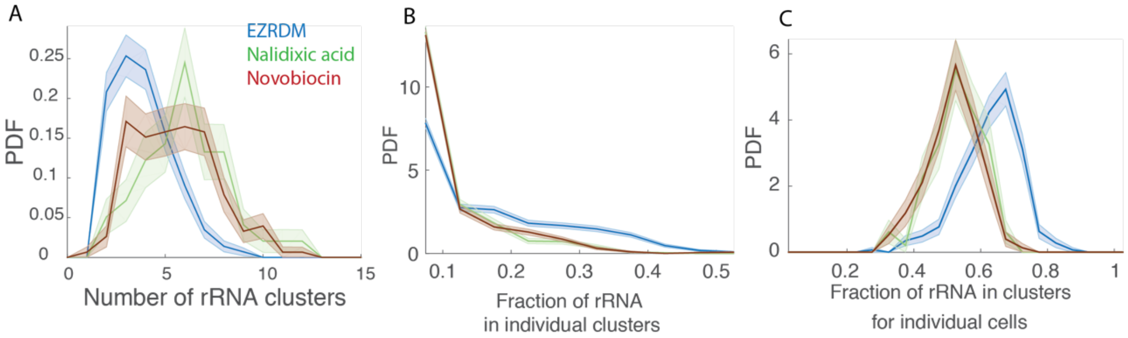


Figure 2.10 Characterization of rRNA clusters under gyrase inhibited conditions. (A) Number of rRNA clusters per cell (B) Fraction of rRNA in individual clusters. (C) Fraction of rRNA in clusters in individual cells. All errors are bootstrapped standard errors.

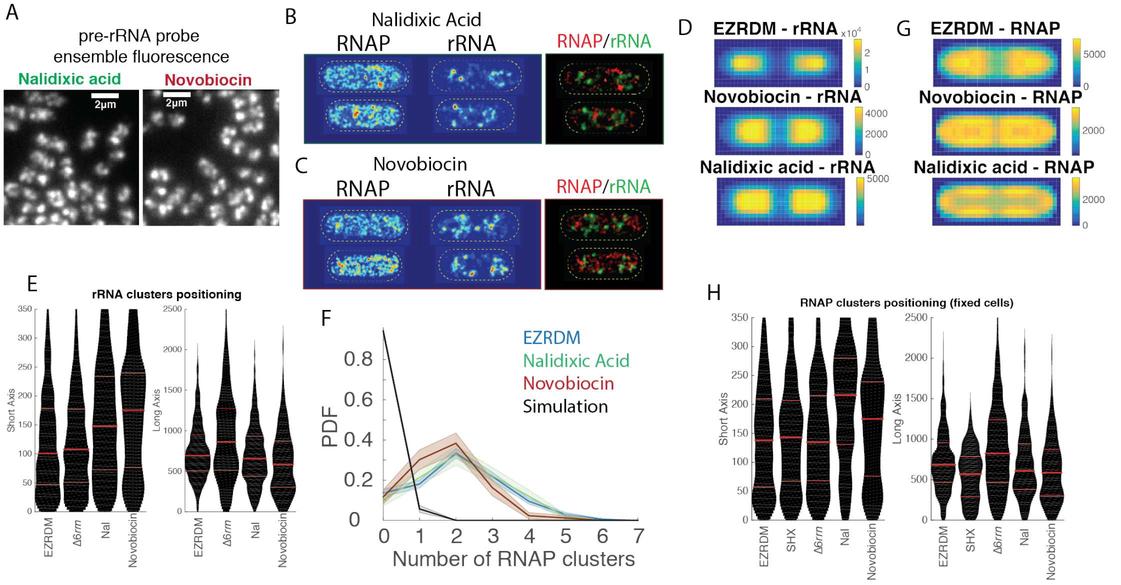


Figure 2.11 Dispersion of RNAP and rRNA clusters in cells under gyrase inhibited conditions. (A) Ensemble fluorescence images showing pre-rRNA signal for nalidixic acid (50ug/mL, 10 min) and novobiocin (300ug/mL, 30 min) treated cells, scale bar is 2μm. Representative superresolution two-color images of RNAP and rRNA in (B) nalidixic acid and (C) novobiocin treated cells, with overlay shown on the right. (D) 2D histogram plots in pseudo-symmetric standard 3x1μm cells showing the distribution of all localizations of rRNA. (E) Cellular positioning of rRNA clusters under various growth conditions, in both the short and long axis; mean, 25th and 75th percentiles are shown in red. (F) Number of RNAP clusters distribution for the EZRDM condition, nalidixic acid and novobiocin treated cells, with negative simulation in black, errors are bootstrapped standard errors. (G) 2D histogram plots of pseudo-symmetric standard 3x1μm cells showing the distribution of all localizations of RNAP. (H) Cellular positioning of RNAP clusters under various growth conditions, in both the short and long axis; mean, 25th and 75th percentiles are shown in red.

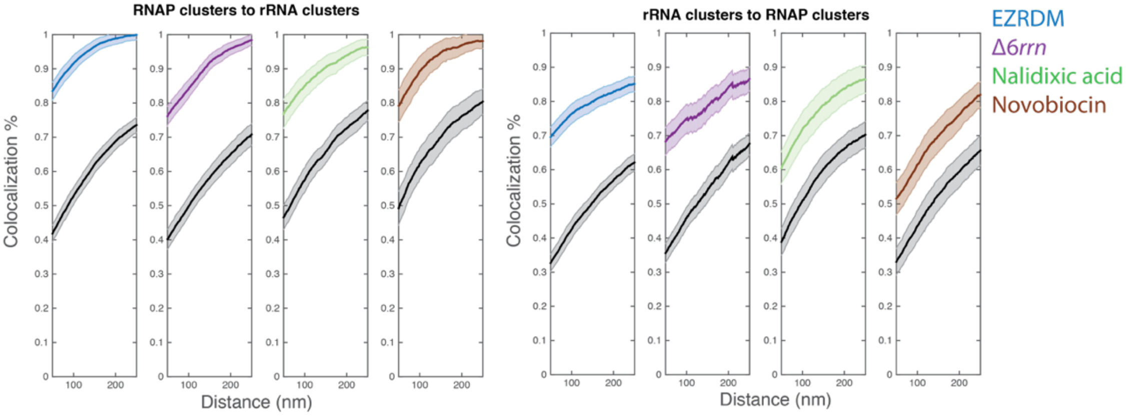


Figure 2.12 Colocalization between RNAP clusters and rRNA cluster for various conditions. Cumulative sum plots for colocalization percentages are shown, negative simulation controls are shown in black. All errors are bootstrapped standard errors.

Cellular positioning of RNAP clusters and *rrn* operons was dependent on cell cycle and transcription activity

Thus far, our results demonstrated that RNAP clusters can exist in the absence of high rRNA synthesis activity, and that a significant component of the colocalization with the *rrnD* operon and the essential transcription factor NusA is independent of transcription activity of *rrn*. These results raised the question of what determines the formation of RNAP clusters, if they are not modulated by transcription activity. Previous immunoprecipitation (IP) experiments have shown that a significant portion of cellular RNAP is bound on chromosomal DNA in promoter or promoter-like regions without active transcription [133]. Thus, it is possible that the underlying spatial organization of chromosomal DNA, orchestrated by specific DNA sequences and likely other protein factors, play an important role in determining the apparent clustering of RNAP. As such, we reasoned that the cellular positioning of RNAP clusters should mimic that of the segregating nucleoids during the cell cycle, as it has been reported previously that individual chromosomal loci follow linearly the movement of the segregating nucleoid [15,19,68-70,73,74,151].

We investigated the cellular positioning of RNA clusters, the *rrnD* DNA site, and NusA clusters. First, we saw that all three species clearly occupy the nucleoid territory when all detected localizations were plotted in standard cell with pseudo-symmetry in a 2D histogram (Figure 2.13Ai-iii). To probe further, we determined centroid positions of individual RNAP clusters (centroid is the mean x and y positions of all localizations within a cluster), and plotted them after binning cells according to cell lengths, which is a common way to infer relative cell ages during the cell cycle [70]. For fast growing cells, if we look at the cell length dependence of short and long axis distribution of RNAP clusters centroids, we saw that RNAP clusters had a fairly consistent axial position through-out the cell cycle, while RNAP clusters moved toward the poles as the cell elongates, implicating a close association of the RNAP clusters to the chromosome (Figure 2.13B). As a comparison, *rrnD* DNA site showed a similar trend of cell length dependent positioning along the long axis (Figure 2.13C), indicating that perhaps RNAP clusters have specific associations to certain sites on the chromosome, which is in turn reflected in the cell length dependent positioning of RNAP clusters. NusA clusters also showed a similar cell cycle dependent positioning along the long axis of the cell (Figure 2.13D); this behavior is not surprising, due to the high level of colocalization between RNAP and NusA clusters. Interestingly, RNAP clusters, *rrnD* and NusA clusters under the other conditions we have tested (SHX, Rif and

Δ6*rrn)*, all showed a similar cell cycle dependent movement toward the poles, indicating that transcription activity or even the presence of the *rrn* operons are not essential for RNAP clusters to associate and move with the nucleoid (Figure 2.14A-F). These observations are consistent with the possibility that the cellular positioning of RNAP clusters is dictated by the underlying chromosomal organization.

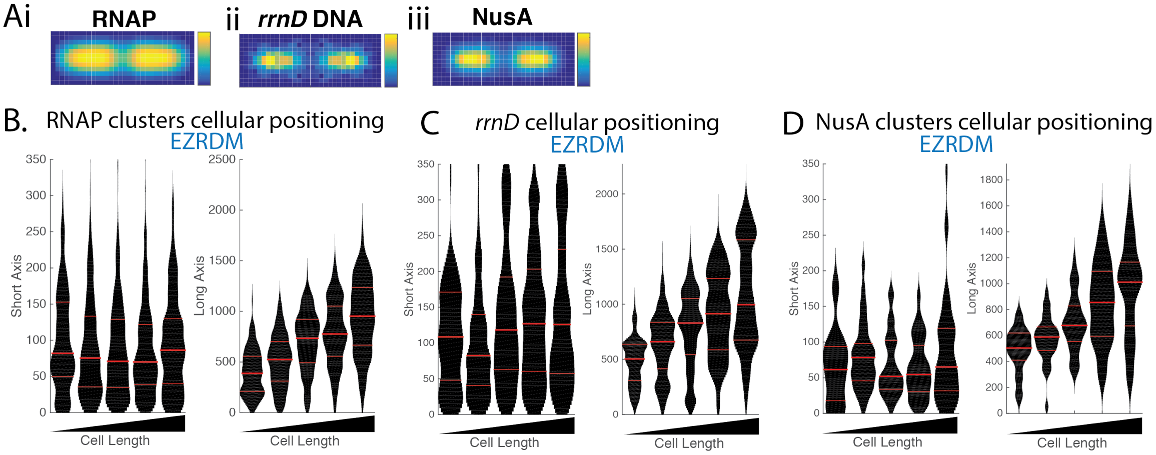


Figure 2.13 Cell-cycle dependent cellular positioning of RNAP clusters, *rrnD* and NusA clusters. (A) Plotting in a 2D histogram in pseudo-symmetry, all localizations for (i) RNAP, (ii) *rrnD* and (iii) NusA in a 3x1 μm standard cell, long axis positions are normalized to cell length. (B) RNAP clusters cellular positioning as a function of cell length in the short and long axis. (C) *rrnD* site's cellular positioning as a function of cell length. (D) NusA clusters cellular positioning as a function of cell length. (B-D) Cells are binned according to length; mean, 25th and 75th percentiles are shown in red, the shape shows the distribution.

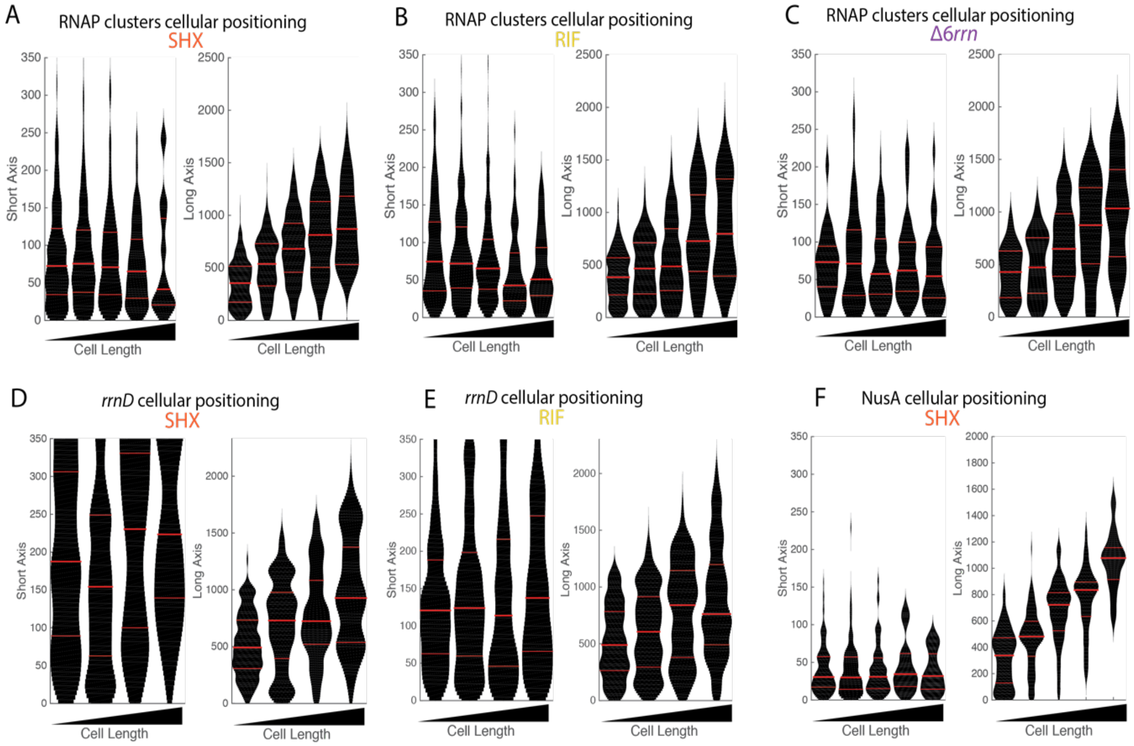


Figure 2.14 Cell-cycle dependent cellular positioning of RNAP clusters, NusA clusters and *rrnD* in various conditions. (A-C) RNAP clusters cellular positioning as a function of cell length in the short and long axis in various conditions. (D, E) *rrnD* site's cellular positioning as a function of cell length in various conditions. (F) NusA clusters cellular positioning as a function of cell length in SHX treated cells. (A-F) Cells are binned according to length; mean, 25th and 75th percentiles are shown in red, the shape shows the distribution.

RNAP cluster colocalization was not specific to the *rrn* operons and had differential colocalization with the individual *rrn* operons

To further investigate what segments of the chromosome are associated with the RNAP clusters, we examined the colocalization between the RNAP clusters and an expanded set of *rrn* operons (*rrnC*, *rrnG*, *rrnH*) and three other mRNA gene sites, including an inducible mRNA gene (*galE*) and two other sites: *mraZ* (as a constitutive promoter region) and *yegH* (to investigate the colocalization of RNAP clusters with the Ter macrodomain) (Figure 2.15A). Previously, we mentioned that for *rrnD*, there is a high level of colocalization between the *rrnD* site and RNAP clusters within 100nm for the fast growth condition (0.52 ± 0.07), and an above-basal level of colocalization for the SHX treated condition (0.33 ± 0.03, Figure 2.15B, C, Table 2.2). Interestingly, none of the other *rrn* operon sites showed this high level of colocalization, though, *rrnC* and *rrnH* were still more significantly colocalized than the basal level of colocalization (Figure 2.15B, C, Table 2.2). The *rrn* operon *rrnG* showed no significant colocalization and was identical to the basal level of colocalization (Figure 2.15B, C, Table 2.2). Surprisingly, *galE* showed a high colocalization value with RNAP clusters, which indicated that colocalization with RNAP clusters was not specific to the *rrn* operons (0.50 ± 0.06 (Figure 2.15B, C, Table 2.2). With galactose induction, *galE* adopted a lower colocalization with RNAP clusters (0.39 ± 0.06), which suggested that perhaps this chromosomal position's spatial relationship to RNAP clusters was not sensitive to the transcription activity of a single mRNA gene (Figure 2.15B, C). The two sites *mraZ* and *yegH* both showed colocalization values higher than that of the *rrnG* operon under the fast growth condition and was more comparable to the *rrnH* and *rrnC* operons (Figure 2.15B, C, Table 2.2). All the colocalization values between the different chromosomal segments and RNAP clusters varied greatly over the other experiment conditions, namely, SHX and Rif.

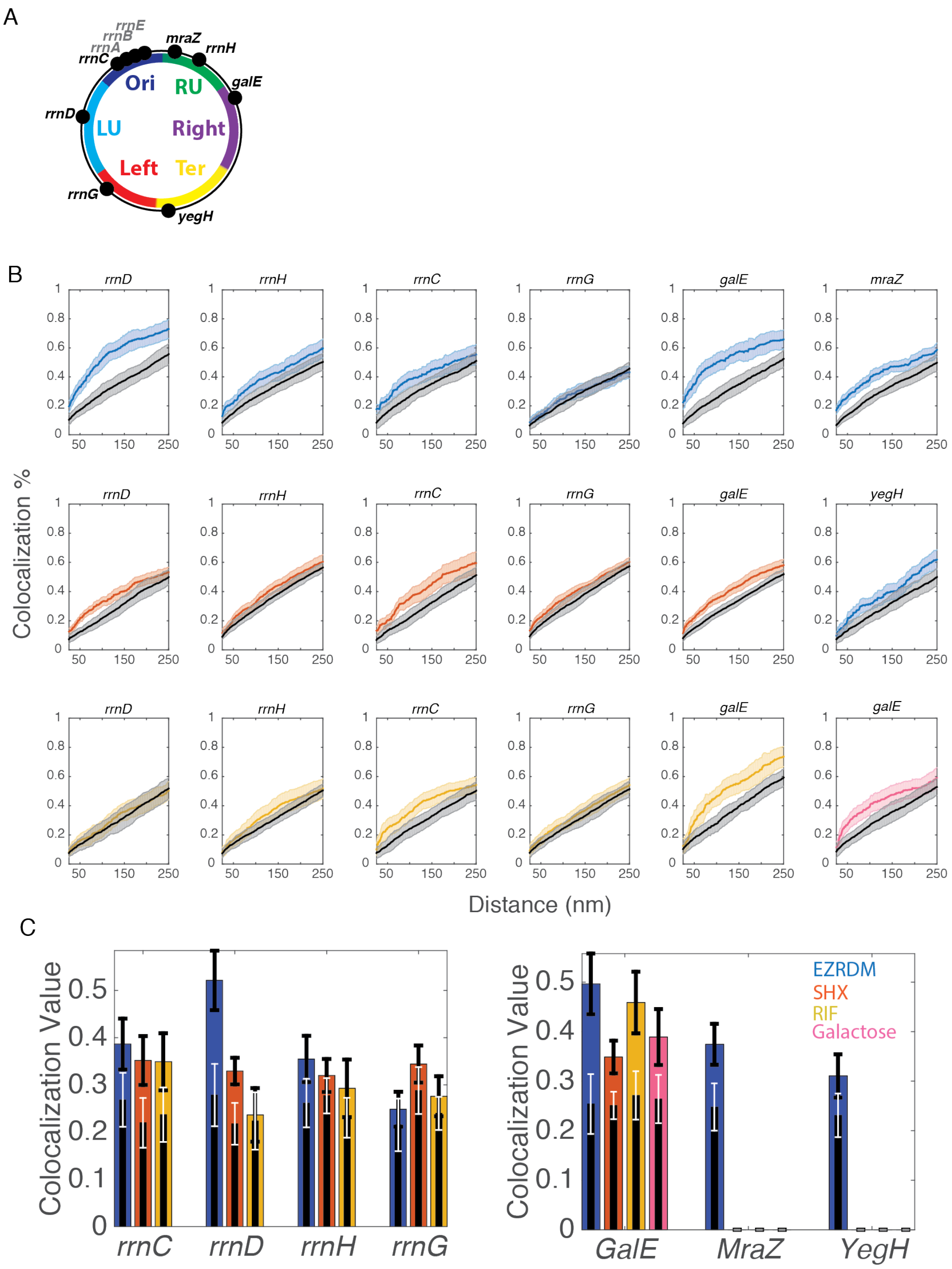


Figure 2.15 Colocalization between operons and RNAP clusters for various conditions. (A) Chromosome map showing positions of singly labeled sites in the strain background of RNAP-PAmCherry. (B) Cumulative sum plots are shown for all DNA-RNAP cluster colocalization strains and conditions; colocalization values reported in text and recorded in Table 2 are based on 100-nm distance threshold. Negative simulation control was run to gauge basal level of colocalization in an experimental RNAP distribution for all conditions, shown in black, all errors are bootstrapped standard errors. (C) Bar plot showing colocalization value using a 100-nm distance threshold, with negative simulation values presented in narrow black bars. All errors are bootstrapped standard errors.

Discussion

Based on these results, we conclude that *E. coli* RNAP is spatially organized to form high density ‘clusters’, where high transcription activity can take place, i.e. for rRNA synthesis during fast growth. This claim is supported by the finding that RNAP clusters during fast growth had the highest proportion of cellular RNAP allocated to these specific regions, likely as the result of recruiting more RNAP molecules during active transcription. The high degree of colocalization between pre-rRNA clusters and RNAP clusters further verified that a large portion of these clusters are associated with each other during fast growth, but under conditions where rRNA synthesis was diminished, we also saw that RNAP clusters can exist independently (Table 2.1 and 2.3). The existence of RNAP clusters without pre-rRNA clusters is clearly seen in Figure 2.5H, suggesting that rRNA transcription may not be necessary for the formation of the RNAP clusters. These results suggest that the formation of RNAP clusters is more likely a reflection of the underlying architecture of the DNA and not due to the transcription of rRNA, similar to the transcription independent tethering of rrn operons and perhaps other DNA sites reported previously [127]. Interestingly, there was a high level of colocalization between RNAP clusters and NusA clusters independent of rRNA synthesis, which suggests that RNAP clusters likely include other protein factors and transcription factors like NusA in an activity-independent manner.

We observed that different *rrn* operons have differential colocalization with RNAP clusters, which suggests that these highly homologous operons could potentially be regulated based on their chromosomal positioning. We also saw RNAP clusters colocalize with non-*rrn* segments of the chromosome, which shows that colocalization with RNAP clusters may not be limited to *rrn* operons, and could include other genes residing on other macrodomains.

Additionally, we showed that RNAP clusters and rRNA synthesis persisted when we perturb the transcriptionally induced supercoiling of the chromosome via gyrase inhibition; both RNAP clusters and rRNA synthesis sites became spatially more dispersed, indicating that activity level may be decoupled from a specific cellular positioning of RNAP clusters. This shift of RNAP clusters and rRNA clusters towards the cell periphery could be a reflection of the changing levels of chromosome compaction due to gyrase inhibition [152].

Materials and Methods

Strain and plasmid construction

The bacterial host strain used for all experiments presented is E. coli K12 MG1655 (Yale Genetic Stock). For detailed strain and plasmid construction information, see (Weng et al. 2018, manuscript in preparation).

Cell growth

Bacterial cells were cultured from single colonies from LB plates in EZ Rich Defined Media (EZRDM, Teknova) with 0.4% glucose, at room temperature (RT) overnight with shaking. Next morning, cells were reinoculated into EZRDM with 0.4% glucose and grown at RT until they reached mid-log phase (OD600 ~0.3-0.4). Antibiotics were added when appropriate for the strain's resistance: 50ug mL -1 kanamycin, 50ug mL -1 carbenicillin. For induction of TetR-EYFP expression, 0.3% L-arabinose was used in EZRDM with 0.4% glycerol, cells were induced for 2hr. For drug treated cells, the time line for TetR-EYFP induction is as follows: 1hr SHX (500ug/mL) treatment was performed during the last hour of TetR-EYFP induction; 2hr RIF (100ug/mL) treatment performed after TetR-EYFP induction. For gyrase inhibition, unless specified it is as follows, nalidixic acid (50ug/mL, 10min), novobiocin (300ug/mL, 30min). Cells were observed immediately after induction or drug treatments.

Sample preparation and imaging

A gel pad made with 3% low-melting-temperature agarose (SeaPlaque, Lonza) in the same growth media (or PBS for fixed cells) was prepared. Cells were spun-down and resuspended in ~50uL of growth media, ~1uL of cells were immobilized between the gel pad and a coverslip for imaging. For fixed cell experiments, cells were fixed in 3.7% (v/v) paraformaldehyde (16% Paraformaldehyde, EM Grade, EMS) for 15min RT, washed with 1X PBS and used immediately. An Olympus IX-81 inverted microscope with a 100X oil objective was used, with 1.6x additional amplification. Images were captured with an Ixon DU-895 (Andor) EM-CCD with a 13-μm pixel width, utilizing the MetaMorph software

(Molecular Devices). Illuminations (405nm, 488nm, 561nm, 647nm) was provided by solid-state lasers as follows: Coherent OBIS-405, Coherent OBIS-488, Coherent Sapphire-561, Coherent OBIS-647. Emitted light was split using a long-pass filter, and the far-red, red and green channels were filtered using HQ705/55, HQ600/50 and ET525/50 bandpass filters (Chroma). For two-color imaging, simultaneous two-color acquisition was achieved using Optosplit II (Cairn Research), colored channels were overlaid using calibration images from TetraSpeck beads (Life Technologies, T-7279), with 10-nm registration error. Gold fiducial beads (50nm, Microspheres-Nanospheres, Mahopac, NY) were used to correct for any sample drift during imaging. All superresolution images were acquired with a 10ms exposure time, for 3000-9000 frames. Activation of fluorescent proteins were done simultaneously to excitation, and kept at a consistent UV power through-out.

Molecule detection in experimental superresolution data

Molecule localization and fitting for all superresolution imaging data were performed using thunderSTORM software (ImageJ) [153]. Subsequent analysis of localizations was performed using custom Matlab routines. For details on fluorescent protein blinking correction, see (Bohrer et al. 2018, manuscript in preparation).

Cluster determination and analysis

To determine a cluster across the different experimental conditions and different molecular species, we first normalized the number of localizations so that each cell had the same concentration of localizations. The concentration normalization eliminates the effects of cell size and the noise in detection efficiency from being the dominating factors in the characteristics of the clusters. To do this we calculated the volume of each cell by outlining the shape of each cell using the outermost localizations to determine an area; this area was then projecting to a 3D volume. We only used cells that had enough localizations to reach a desired concentration threshold. Second, we eliminated localizations in low-density areas. To do this we calculated the average distance to the closest ten localizations surrounding the localization of interest, and if this was greater than a specified value, the localization was not utilized in the third step. Third, we used the tree cluster analysis software commercially available in MATLAB. Explicitly, we utilized the 'single' method using the linkage function, which provided us with a tree of hierarchical clusters for the data. We then used the cluster function with a cutoff of 100nm and the distance criterion. This analysis links all localizations together as one cluster if they are within 100nm of each other. As a final step, we only counted clusters that possessed more than a certain percentage of the total localizations in the cell.

Colocalization value calculation

We calculated the colocalization value between two different 'species' by doing the following: a colocalization value is always defined from species 1 to species 2, as the reverse colocalization value is not the same; the closest distance between any of the localizations that are in species 1 to any of the localizations in species 2 is calculated for every species 1. There must be at least one of each species in a cell to contribute a measurement to the colocalization value. The colocalization value at a particular distance was the percentage of distances that are smaller than that particular distance. The colocalization definition was the same for every species investigated except for the superresolution imaging of the DNA locations. For the DNA locations, all localizations within 50nm of each other are linked together, the mean of all the linked localizations was then utilized in the calculation of the colocalization value. We sometimes display the colocalization between two species as a cumulative distribution to show the colocalization value for each distance, alternatively, we can report a particular colocalization value at a set distance threshold for ease of comparison between conditions.

Adjusting colocalization values for experimental detection efficiency

In this work, all colocalization values and the colocalization cumulative distributions were adjusted for the detection efficiency of the second species (from species 1 to species 2). To determine the detection efficiency of a particular species, i.e. rRNA clusters, RNAP clusters, or NusA clusters, we utilized the fact that the number of localizations was adjusted to a predefined concentration in the first step of the cluster detection algorithm, see Cluster Determination. First, only the cells that have at least twice the predefined concentration of localizations were used in the calculation of the detection efficiency. We then randomly split the localizations of these cells so that there were two sets of localizations of the same molecule in a cell with the desired concentration. Second, we performed the cluster analysis on each set of localizations and then calculated the colocalization between them. The colocalization cumulative distribution of this analyses provided us with the best possible colocalization at each distance given our detection efficiency if species 1 is smaller than or equal to species two. To verify this approach, we performed a control experiment: calculated rRNA cluster detection efficiency using two L1 probes with different colored dyes attached (Figure 2.16). The detection efficiency calculated by this approach agreed with the detection efficiency calculated from our methodology.

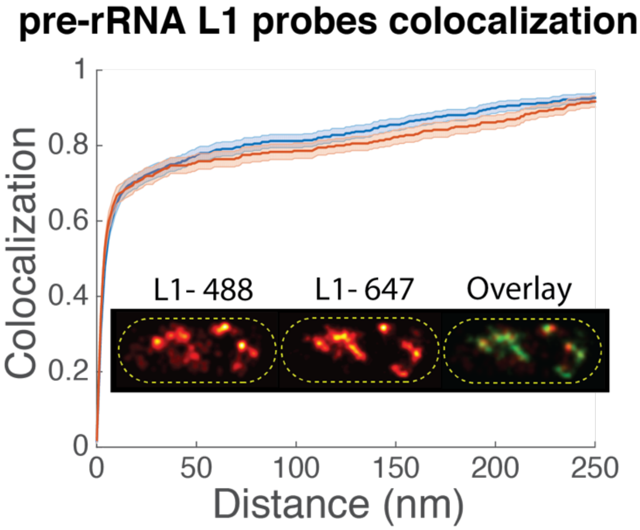


Figure 2.16 rRNA cluster detection efficiency measurement with L1-L1 probes. Two L1 probes were used to measure the detection efficiency of pre-rRNA clusters, L1-488 and L1-647 have the same sequence but different dye labels. Two-color superresolution imaging were performed on cells hybridized with both probes (representative cell image shown in inset). The cumulative curves of the percentage of one probe colocalizing to the other (red and blue curves) show high detection efficiencies of pre-rRNA clusters using either probe. The detection efficiency was estimated to be ~90% for either probe at the distance threshold of 50-nm.

Negative simulations for colocalization values

The negative simulations in this work were used to determine the amount of colocalization to the first species (species 1), in a colocalization calculation, coordinates inside of the cell followed a uniform distribution. We calculated the volume of each cell by outlining the shape of each cell using the outermost localizations to determine an area; this area was then projecting to a 3D volume. We then randomly distributed the number of localizations as the species within this volume. If species 1 was RNAP, NusA or rRNA, we added 100 more localizations around each of the previously simulated localizations according to a normal distribution with a std = 50 nm. Simulating clusters of molecules was done to account for the width contributions of species 1 being a cluster itself.

Pre-rRNA single-molecule fluorescence *in situ* hybridization (smFISH)

smFISH experimental protocol closely followed a previously described protocol [154]. Briefly, cells were grown in EZRDM glucose as previously described, 5 mL of mid-log phase cells were fixed with 3.7% (v/v) paraformaldehyde (16% Paraformaldehyde, EM Grade, EMS), placed for 30 min on ice. Next, cells were harvested via centrifugation, and subsequently washed 2X in 1X PBS. Cells were then permeabilized by resuspension in 300 uL of H_2O followed by 700 μL of 100% ethanol, mix well and rotate at RT for 30 min. At this stage cells could be stored at 4°C until next day. Freshly prepare wash buffer with 40% formamide and 2x SSC and put on ice. Spin-down cells and resuspend in 1mL of wash buffer, mix for 5 min at RT. Prepare hybridization solution with 40% formamide and 2x SSC, add oligo probes to final concentration of 1 μM. Spin down cells in wash buffer and add 50uL of hybridization solution with probe for each sample, mix and let incubate O/N at 30°C. Next day, incubate 10 μL of hybridization sample with 200 μL of fresh wash buffer at 30°C for 30 min, 2X. The rest of the hybridization sample could be stored at 4°C for a week. The washed sample was imaged immediately, without STORM imaging buffer for ensemble fluorescence, with STORM buffer to induce dye blinking for superresolution imaging. glucose oxidase + thiol STORM buffer was used to image samples with only dye-labeling (50 mM Tris (pH 8.0), 10mM NaCl, 0.5 mg ml^{-1} glucose oxidase (Sigma-Aldrich), 40ug ml^{-1} catalase (Roche), 10% (w/v) glucose and 10mM MEA (Fluka)) [155]. Thiol only STORM buffer (10mM

MEA, 50 mM Tris (pH 8.0), 10mM NaCl) was used to image samples with both endogenously expressed fluorescent proteins and dye-labeling, to preserve fluorescent signal from fluorescent proteins.

Pre-rRNA transcripts were labeled with a single probe called L1, conjugated at the 5' with either Alexa-488 (NHS ester) or Alexa-647(NHS ester) (IDT) [132]. 50uM stocks were made and aliquoted upon receiving the commercial oligos, these were stored protected from light.

ImmunoSTORM of RNAP by labeling β and β'

Fixation was with 4% (v/v) paraformaldehyde and 0.05% (v/v) glutaraldehyde for 30 min on ice. Samples were prepared post fixation following a standard immunostaining protocol. Briefly, primary antibodies are RNAP mouse α-RNAP β, and mouse α-RNAP β', used at 1:500 dilutions. Secondary antibodies, Alexa Fluor 647 goat α-mouse (Life Technologies, A-11001) were used at 1:500 dilutions. After labeling procedure, cells were imaged immediately with glucose oxidase + thiol STORM buffer.

Image processing of ensemble fluoresence images (pre-rRNA)

Ensemble intensity measurements were performed using ImageJ. Ensemble fluoresence images with focus around mid-cell were segmented manually using their corresponding bright-field images as a guide, each cell's total fluorescent intensity was calculated as: (area of the segmented cell x (mean intensity inside cell – mean intensity of background region)). Around 100 cells were used to represent the total cellular fluoresence intensities for a single condition.

Chapter 3: Visualization of DNA Loops in Live *E. coli*

Introduction

Looping between two DNA sites, mediated by transcription factors, is a ubiquitous mechanism in prokaryotic transcription regulation [156]. DNA looping brings two distal DNA sites into close proximity, enhancing interactions between transcription factors bound at separate sites or bringing transcription factors close to RNA polymerase at the promoter. Knowing when and how DNA loops form *in vivo* is important to understand the role of DNA looping in gene regulation and cell decision-making; some studies found molecular details of gene regulation have little influence on gene expression [157-159], while others suggested that DNA looping could trigger cell phenotype switching [160] and influence fluctuations in transcription activity [161].

DNA looping was first suggested for the transcription factor AraC in the *E. coli* arabinose operon. Disruption of an AraC binding site ~280 bp upstream of the promoter reduced AraC-mediated repression nearly ten-fold, indicating a long-range interaction between the promoter and upstream DNA [162]. Subsequently, DNA looping mediated by transcription factors LacI [163], DeoR [164], NtrC [165], GalR [84] and bacteriophage λ repressor CI [166,167] was reported. The length of the intervening DNA in these loops can be as short as 58 bp (*lac* operon [163]) or as long as ~5 kilobases (*deo* operon [164]).

Biochemical, biophysical and genetic studies have established important roles of DNA looping in transcription regulation. However, transcription-factor-mediated DNA looping on the length scale of a few kilobases in prokaryotic cells has not been directly visualized *in vivo*, and the *in vivo* dynamics of DNA looping are difficult to investigate. Chromosome conformation capture (3C) has been used to detect juxtaposition of DNA sites separated by hundreds of kilobases in both eukaryotic and prokaryotic cells [28,168], but high background of interactions at the kilobase scale limits the utility of these methods in studying typical prokaryotic DNA loops [169]. An *in vivo* imaging method using fluorescent proteins fused to DNA-binding proteins bound to tandem arrays of hundreds of binding sites has been employed to visualize homologous chromosome pairing in yeast induced by double-strand breaks [170]; however, an array of several kilobases of binding sites makes this method unsuitable for studying DNA loops of only a few kilobases. In addition, the long array of tightly bound protein molecules may be detrimental to cells [171].

We developed a two-color, high-resolution imaging method to directly measure the end-to-end separation of two DNA sites 2.3 kb apart in live *E. coli* cells (Figure 3.1A). This method is based on the ability to precisely determine the location of a specific DNA site *in vivo* [60]. By expressing a fluorescent protein in fusion with a DNA-binding protein in a cell with only three tandem binding sites (spanning less than 100 bp), the resulting fluorescent spot is diffraction-limited, and the location of the binding site can be determined with sub-diffraction-limited precision by fitting its fluorescence profile to a two-dimensional Gaussian function [77]. By labeling two ends of a DNA segment with two unique sets of binding sequences and co-expressing corresponding fluorescent DNA-binding fusion proteins of different colors, the distance between the two DNA sites can be determined with a precision of a few tens nanometers. An *in vitro* experiment employing the same principle measured intramolecular distances using organic dyes [172], but this approach has not been demonstrated *in vivo* with comparable resolution using fluorescent proteins.

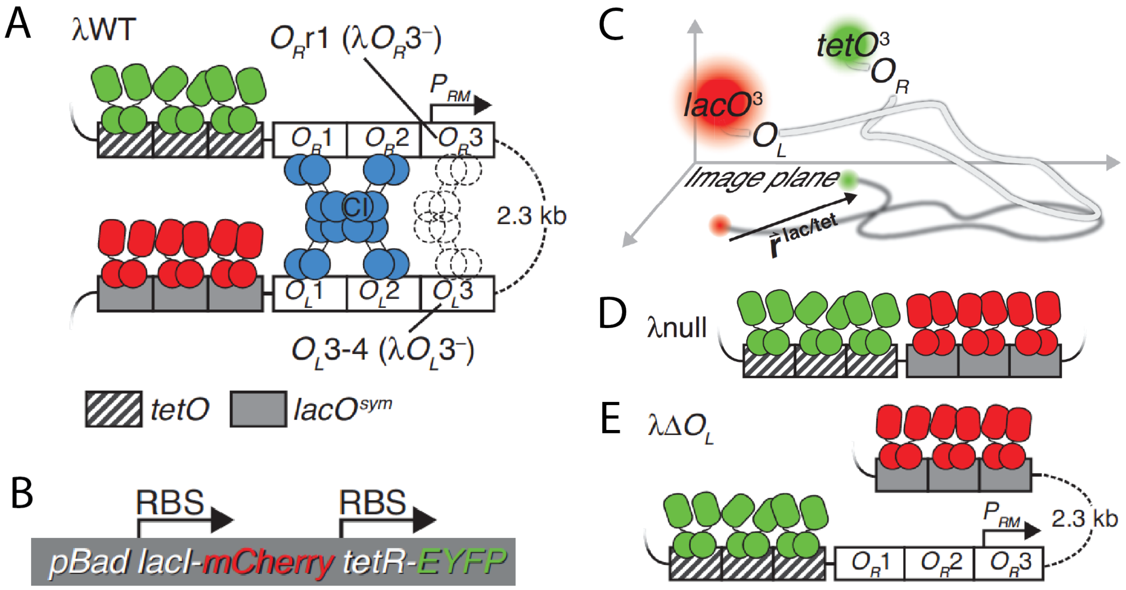


Figure 3.1 Visualizing DNA looping *in vivo* by localizing O_R and O_L with fluorescent DNA-binding fusion proteins. (A) λWT construct. Three tandem *lacO*^{sym} and *tetO* sites, termed *lacO*^{3} and *tetO*^{3}, were placed immediately next to O_L and O_R, respectively. Red and yellow fluorescent fusion proteins LacI-mCherry and TetR-EYFP bind *lacO*^{3} and *tetO*^{3}, respectively. DNA looping mediated by a CI octamer (blue) or an additional CI tetramer (dashed) brings *lacO*^{3} and *tetO*^{3} together. Strains λO_R3^- and λO_L3^- harbor mutations (described in main text) to O_R3 and O_L3, respectively, that prevent CI dimers from binding these operator sites. (B) LacI-mCherry and TetR-EYFP are expressed co-transcriptionally from separate ribosome binding sites on a plasmid. (C) Illustration of $\vec{r}^{\,\mathbf{lac/tet}}$ measurement. The observed distance between mCherry and EYFP spots indicates the distance between *lacO*^{3} and *tetO*^{3} projected onto the imaging plane. (D) Positive control λnull. The centers of *lacO*^{3} and *tetO*^{3} are separated by only 66 bp. (E) Negative control λΔO_L. O_L is deleted to eliminate CI-mediated DNA looping.

We used our method to probe the mechanisms and dynamics of DNA looping mediated by the bacteriophage λ repressor CI in live *E. coli* cells and investigate its regulation of transcription from the CI promoter P_{RM}. The λ repressor CI is an essential transcription factor in determining the fate of an *E. coli* cell infected by the bacteriophage λ. When CI is expressed, it represses lytic promoters to commit to an extraordinarily stable lysogenic state that persists for millions of generations [173-175]. However, upon induction by UV irradiation or other specific events, CI degradation can trigger an irreversible switch from lysogenic to lytic gene expression within one cell generation time [176].

The robustness of the λ regulatory circuit has been extensively studied. Among many important features of the system-such as promoter-operator arrangement [177,178], CI autoregulation [158,179,180] and cooperative binding [181-184]-DNA looping between the homologous rightward and leftward operators O_R and O_L, separated by 2.3 kb, was shown to play significant, fate-determining roles in the λ lifecycle [167,185]. The cooperative binding of CI dimers at the subsites O_R1 and O_R2 of O_R represses the lytic promoter P_R and simultaneously activates CI's own promoter, P_{RM}, by accelerating transcription initiation [186-188]. At higher CI concentrations, an additional CI dimer binds to O_R3 and represses P_{RM}.

As illustrated in Figure 3.1A, an octameric CI complex (with or without an additional CI tetramer) can mediate DNA looping by bridging O_R and O_L. These higher-order complexes result from interactions between CI dimers bound to subsites at O_R123 and O_L123, and were first identified *in vitro* by ultracentrifugation and later visualized by EM [166] and AFM [189]. Looping dynamics were investigated *in vitro* using tethered particle motion (TPM) [190-193].

In this study, we tracked the apparent separation between the O_R and O_L sites on a λ DNA segment (termed O_R–O_L DNA below) in real time in live *E. coli* cells, from which we obtained the first direct estimates of *in vivo* looping frequencies for both wild-type DNA and for DNA carrying mutations in O_R3 and O_L3. Furthermore, we discuss how the compaction of the *E. coli* chromosome may impact DNA looping. The methodology established in this work can be extended to a broad range of questions regarding chromosomal DNA conformation and/or gene activities in prokaryotes and higher organisms.

Results

High-resolution imaging of two DNA sites

We inserted the construct shown in Figure 3.1A into the *E. coli* chromosome. It contains three tandem *tetO* sites (*tetO*^{3}) and three tandem *lacO*^{sym} sites (*lacO*^{3}) [194] flanking the wild-type λ lysogen sequence from O_R to O_L (including the P_R, P_{RM} *and* P_L promoters and the *cI*, *rexA* and *rexB* genes). In this λWT construct, CI is expressed from P_{RM} and regulates its own expression. The *lacO*-binding and *tetO*-binding proteins LacI and TetR were fused with red and yellow fluorescent proteins to generate LacI-mCherry and TetR-EYFP, and were expressed from an inducible plasmid (Figure 3.1B). With the combination of strong induction, weak ribosome binding sites and carefully controlled growth, we achieved sufficiently low LacI-mCherry and TetR-EYFP expression levels to detect distinct, diffraction-limited mCherry and EYFP spots in single cells. We then fit the fluorescence intensity profile of each individual spot with a two-dimensional Gaussian function to estimate its centroid position. The average localization precisions for individual spots of LacI-mCherry and TetR-EYFP were 17 and 14 nm, respectively (Figure 3.2A). Subsequently, we transformed EYFP coordinates into mCherry coordinates using fiducial data to calculate the vector between the mCherry and EYFP spots arising from LacI-mCherry and TetR-EYFP protein molecules bound to the same O_R–O_L DNA segment. We called this vector $\vec{\boldsymbol{r}}^{\mathbf{lac/tet}}$ (Figure 3.1C). The magnitude of the vector, $r^{\mathrm{lac/tet}}$, is the two-dimensional projection of the distance between *lacO*^{3} and *tetO*^{3} onto the image plane; on average, it is proportional to the end-to-end distance between *lacO*^{3} and *tetO*^{3} in three dimensions. The total error for an $r^{\mathrm{lac/tet}}$ measurement, including fitting errors in determining centroid of individual spots (Figure 3.2A), registration errors in aligning EYFP and mCherry two-color images (~10 nm based upon experiments using fluorescent beads) and contributions from local fluorescent background, was on average ~40 nm (see below). With very low TetR-EYFP and LacI-mCherry expression, not all *lacO*^{3} and *tetO*^{3} sites were bound by fusion protein molecules. Furthermore, not all fusion protein molecules were fluorescent due to stochastic chromophore maturation. We analyzed all cells having distinct fluorescent spots in both emission channels to calculate $\vec{\boldsymbol{r}}^{\mathbf{lac/tet}}$. We expected $r^{\mathrm{lac/tet}}$ to decrease when DNA between *lacO*^{3} and *tetO*^{3} looped.

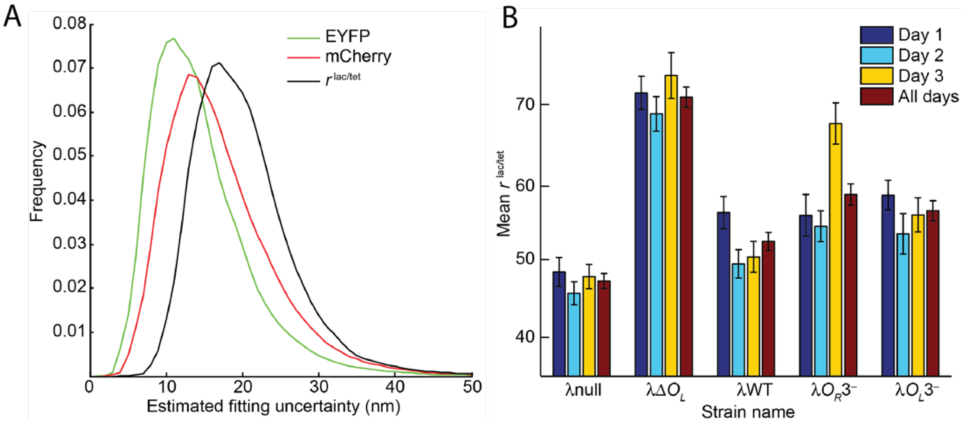


Figure 3.2 Spot fitting and experimental error analysis. (A) Distribution of fitting errors for EYFP (green), mCherry (red) localizations and $r^{\text{lac/tet}}$ (black). Errors were estimated using a bootstrapping procedure by fitting raw data to a Gaussian distribution. The residuals from this fit were then randomly rearranged and added back to the data in 10 different permutations. The reported error is the standard deviation of the distance between these 10 locations and the initial fit location. Error in $r^{\text{lac/tet}}$ was determined similarly; from the 10 bootstrapped EYFP and mCherry fits, 100 distances were obtained and the error was estimated as the standard deviation of the difference between these distances and the distance determined from fitting the raw data. (B) A compilation of all data from three separate experiments was used for all analysis in the main text. Here, $\langle r^{\text{lac/tet}} \rangle$ is shown for the individual experiments. Error was estimated as the standard deviation of the means of 1,000 bootstrapped distributions. Except for one sample (λO_R3^-, day 3), the estimated mean separations for all days followed the trend λnull < λWT < λO_R3^- ≃ λO_L3^- < λΔO_L^-.

Distinguishing between looped and unlooped states

To determine whether our two-color imaging method was sufficient to distinguish between looped and unlooped DNA in the crowded intracellular environment, we constructed two control strains (Table 3.1). In the positive control λnull, the centers of *lacO*^{3} and *tetO*^{3} *sites* are separated by 66 bp **(**Figure 3.1D). The outmost *lacO*^{sym} and *tetO* sites are separated by less than 40 nm (Figure 3.3A). The close proximity of *lacO*^{3} and *tetO*^{3} mimicked permanently looped DNA. In the negative control λΔO_L, we inserted the λ sequence from O_R up to but not including O_L between *lacO*^{3} and *tetO*^{3} (Figure 3.1E). The resulting λΔO_L DNA has comparable length as the wild-type λ DNA, but CI-mediated DNA looping between O_R and O_L is abolished.

Table 3.1 Descriptions of strains and plasmids used in this study. *The 2.3-kb, wild-type phage λ sequence from O_R to O_L was incorporated into the E. coli chromosome in λWT. Strains λO_R3^-, λO_L3^- and λWT^{G147D} contain the r1, OL3-4 and CI^{G147D} mutations, respectively. Curly brackets indicate fusion products expressed from a single ribosome binding site.*

Strain Name	Genotype
λnull	MG1655 Δ*lacIzya*::$lacO^3tetO^3$
λWT	λnull $lacO^3tetO^3$::(O_R–O_L)
λΔO_L	λWT ΔOL
λO_R3^-	λWT O_R3r1[79]
λO_L3^-	λWT O_L3-4[35]
λΔ$O_LP_{RM}^-cI^-$	**λΔO_L $P_{RM}^-cI^-$**
λΔ$O_LP_{RM}^-cI^-/cI^{trans}$	**λΔO_L $P_{RM}^-cI^-$ (pACL18-cI^{wt})**
λWT^{G147D}	**λWT *cI(G147D)* [62]**
pZH102R33Y29	pLau53[81] *pBad*-{*lacI-mCherry*}-{*tetR-EYFP*}
pZH102R33TD	pLau53 *pBad*-{*lacI-mCherry-EYFP*}
pACL18	**pACYC184-cI^{wt}**

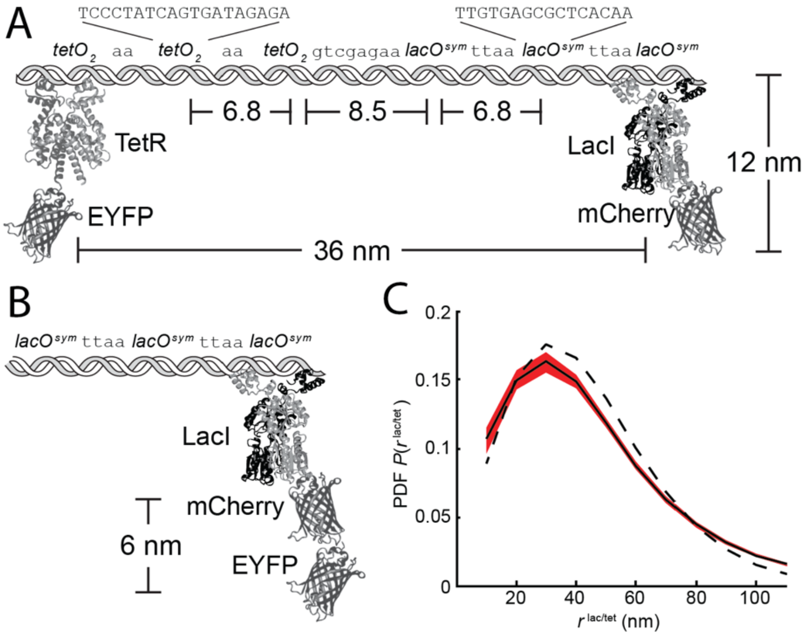


Figure 3.3 Estimate of positive control dimensions and apparent end-to-end distance distribution. (A) The maximum distance between TetR-EYFP and mCherry-LacI chromophores was approximated assuming straight DNA. All distances are in nm. Here, bound fusion proteins are shown on the same face of a DNA molecule, but this needs not be the case. Dimers of DNA-binding proteins were based on Protein Data Bank (PDB) entries for TetR (1QPI) and LacI (1EFA). Both fluorescent proteins are shown using the entry for GFP (1GFL). Protein structures images generated using VMD. (B) In an alternative positive control that was used to collect fiducial data for image registration, the plasmid pZH102R33TD encodes the tandem-dimer reporter LacI-mCherry-EYFP. (C) The $r^{\mathrm{lac/tet}}$ PDF for the λnull control (black line; 1 s.e.m. shown in red as in Fig. 3A) is shown with the distribution of 10,000 numerically simulated end-to-end distances for two sites separated by 22-nm, randomly projected onto the 2D plane, and subjected to 22-nm localization error for both ends (dashed black line). PDFs were calculated using methods described in main text.

We first examined λnull and λΔO_L in two-color fluorescence images to determine whether we could discriminate between looped and unlooped DNA by eye. We obtained at least sixty 20-frame movies (100 ms exposures; 2 s total) for each strain in each of three independent experiments. Typical fluorescence images are shown in Figure 3.4A and B. Crosstalk between the two emission channels was negligible, as bright mCherry and EYFP spots only appeared in the corresponding channel but not the other.

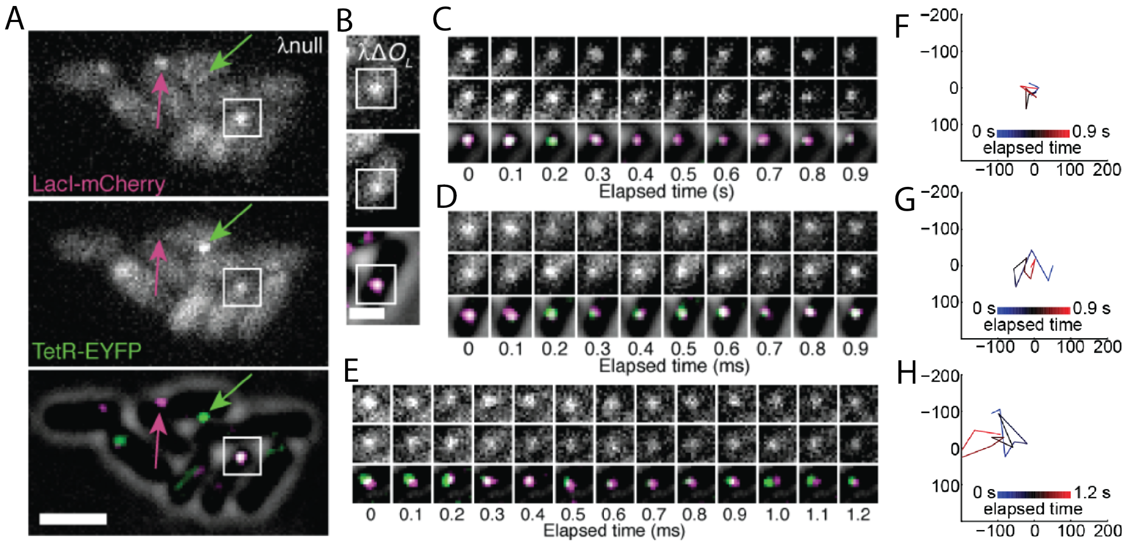


Figure 3.4 High-resolution imaging of *lacO*^{3} and *tetO*^{3} sites separated by 66 bp (λnull) or 2.3 kb (λΔO_L). (A) Fluorescent images of λnull. Arrows highlight molecules that exclusively appeared in mCherry (magenta, top) and EYFP (green, middle) channels, indicating a lack of significant crosstalk between the two channels. Squares show a spot that appeared in both channels. In the overlay image (bottom), fluorescence images were bandpass filtered and background was subtracted. Only cells having both mCherry and EYFP fluorescence were used in analysis. Scale bar 2 μm. The image order and color scheme are repeated in B–E. (B) Fluorescent images of λΔO_L. Scale bar 1 μm. (C–E) Timelapse images of spots acquired every 100 ms; C and D are spots in white squares in A and B, respectively, and E shows an additional λΔO_L spot, whose apparent separation can be easily detected by eye. Top and middle rows show mCherry and EYFP channels, respectively, and bottom rows show two-color overlays on brightfield images. (F–H) Trajectories $\vec{r}^{\,\mathbf{lac/tet}}$ vectors from fitting fluorescence data for spots in C–E. Coordinates are in nm. Vertices indicate the $\vec{r}^{\,\mathbf{lac/tet}}$ vector and subsequent time points are connected by lines that are colored to indicate elapsed time.

Figure 3.4C and D show 1 s of typical data for individual λnull and λΔO_L spots. As expected for a permanently looped configuration, the positive control λnull exhibited overlapping EYFP and mCherry spots (Figure 3.4C). Most λΔO_L molecules also did not exhibit clear spot separation that was easily identifiable by eye (Figure 3.4D). However, some λΔO_L molecules displayed large displacements between the LacI-mCherry and TetR-EYFP spots that were distinguishable by eye (Figure 3.4E); such images were not observed for λnull.

Visual inspection of the apparent separation between the LacI-mCherry and TetR-EYFP spots suggested that comparing the end-to-end separation in O_R–O_L DNAs required a more quantitative approach. We calculated $\vec{\boldsymbol{r}}^{\mathbf{lac/tet}}$ for all O_R–O_L DNA molecules in the λΔO_L and λnull strains that exhibited fluorescent spots in both EYFP and mCherry images. Figure 3.4F-H show $\vec{\boldsymbol{r}}^{\mathbf{lac/tet}}$ calculations for movies in Figure 3.4C-E, respectively. We then compiled the corresponding probability density distributions (PDF, $P\left(r^{\mathrm{lac/tet}}\right)$, Figure 3.5A) and cumulative density distributions (CDF, $C\left(r^{\mathrm{lac/tet}}\right)$, Figure 3.5B) of the vector magnitude, $r^{\mathrm{lac/tet}}$. The long-tailed PDF observed for λnull (Figure 3.5A) is consistent with the expected end-to-end distance distribution measured from two spots with a fixed separation when the localization of each spot is subject to Gaussian fitting error [195].

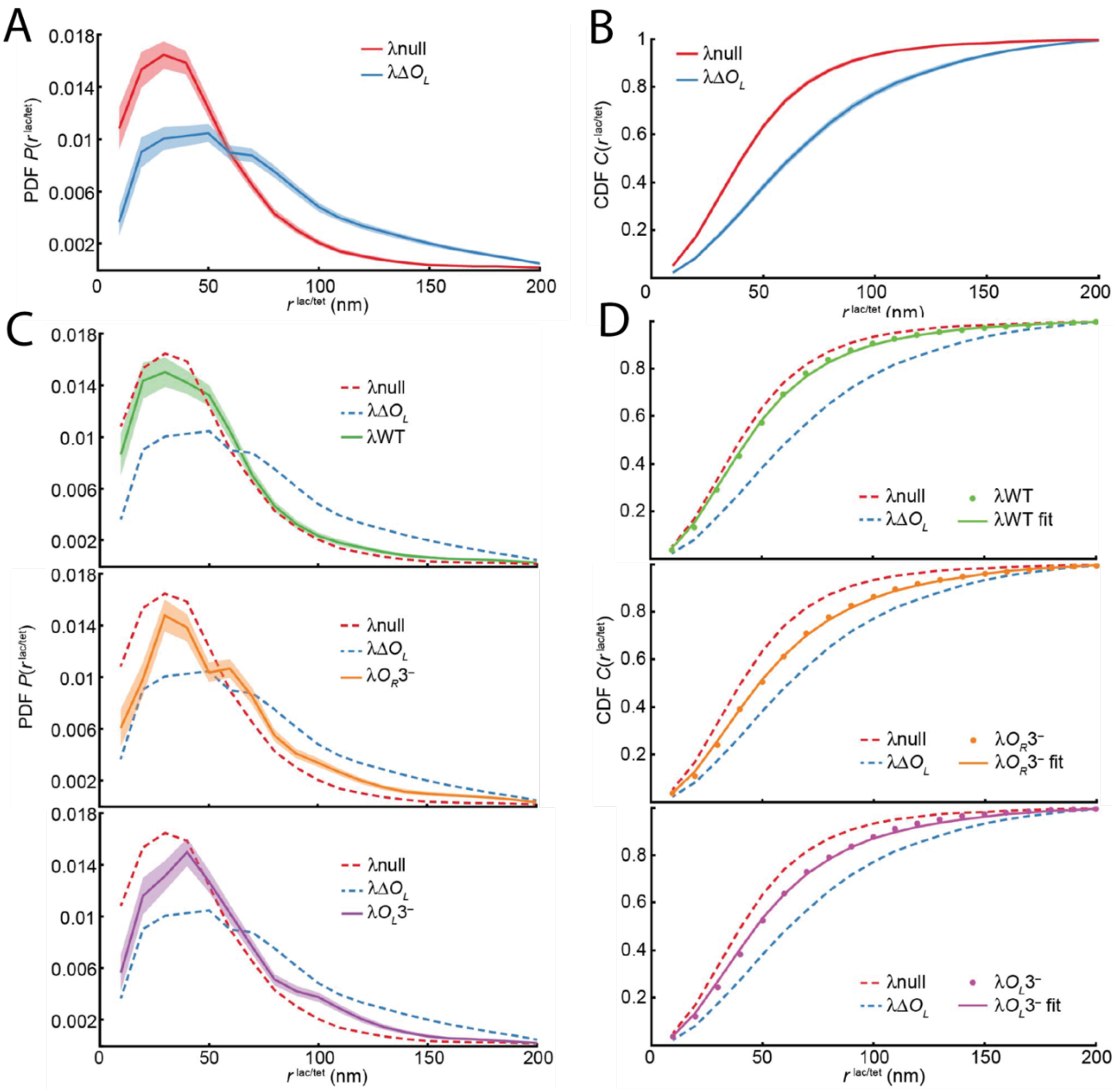


Figure 3.5 End-to-end distance ($r^{\text{lac/tet}}$) distributions and looping frequency fitting. (A) Probability density distribution (PDF) of the $\vec{r}^{\,\text{lac/tet}}$ vector magnitude $r^{\text{lac/tet}}$ for the looped (λnull, red) and unlooped (λΔ$O_{L,}$ blue) controls. The PDF is estimated for 10-nm bins as described in the main text. Light-colored areas indicate 1 s.e.m. calculated by bootstrapping. (B) Cumulative density (CDF) of $r^{\text{lac/tet}}$ for the looped (λnull) and unlooped (λΔO_L) controls. The CDF is estimated for 10-nm bins as described in the main text. Light-colored area indicates 1 s.e.m. calculated by bootstrapping. (C) The PDF is shown for strains λWT (green), λO_R3⁻ (orange) and λO_L3⁻ (purple), calculated as in A, and PDFs for strains λnull and λΔO_L are shown as dashed lines for comparison. (D) CDF estimates for three strains (dots; λWT, green; λO_R3⁻, orange; λO_L3⁻, purple) were fit as linear combinations of the positive (λnull) and negative (λΔO_L) control CDFs to estimate looping frequency. Colored lines indicate CDF fits and CDFs for strains λnull and λΔO_L are shown as dashed lines for comparison.

We found that the $r^{\text{lac/tet}}$ distribution for λΔO_L was distinctly different from that of λnull ($p < 10^{-3}$); the difference was reproduced in three independent experiments (Figure 3.2B). The mean separations, $\langle r^{\text{lac/tet}} \rangle$, were 47 (N=1153) and 71 nm (N=979) for λnull and λΔO_L respectively (results and measurement errors summarized in Table 3.2). Peaks in $P(r^{\text{lac/tet}})$ plots centered at ~40 nm, reflecting our experimental precision in determining $r^{\text{lac/tet}}$; *i.e.* O_R–O_L molecules with $r^{\text{lac/tet}}$ below 40 nm could not be distinguished from each other. Hence, it was more meaningful to compare distributions of $r^{\text{lac/tet}}$ at large values where $r^{\text{lac/tet}}$ distributions differed most prominently. The cumulative probability of $r^{\text{lac/tet}}$ being 75 nm or more was ~40% for λΔ*OL* and only ~15% for λnull (Figure 3.5B**)**. By examining $r^{\text{lac/tet}}$ distributions, we could distinguish between the looped and unlooped control strains, suggesting that this approach could be used to probe CI-mediated DNA looping.

Table 3.2 Summary of measurements and sample statistics in this study. *Errors are all 1 s.e.m. as estimated from 1000 bootstrapped samples.*

Strain	$r^{\text{lac/tet}}$ measurements	Mean $r^{\text{lac/tet}}$ (nm)	Median $r^{\text{lac/tet}}$ (nm)	Looping frequency	CI expression level (WLU)
λnull	1153	47 ± 1	41 ± 1	N/A	N/A
λΔO_L	979	71 ± 1	63 ± 2	N/A	1.38 ± 0.05
λWT	962	52 ± 1	45 ± 1	79 ± 6%	1.00 ± 0.05
λO_R3^-	784	59 ± 1	50 ± 2	53 ± 7%	2.50 ± 0.07
λO_L3^-	781	56 ± 1	48 ± 1	60 ± 7%	2.51 ± 0.07

Compact conformation of unlooped DNA λΔO_L does not depend on transcription or non-specific CI binding

We measured the mean end-to-end separation $\langle r^{\text{lac/tet}} \rangle$ for λΔO_L at 71-nm, much shorter than the ~200-nm distance expected for B-form DNA with a typical

50-nm *in vitro* persistence length [196]. While such a result is expected given the many factors known to compact prokaryotic chromosomes [197], it is possible that non-specifically-bound CI on the $\lambda\Delta O_L$ DNA and/or P_{RM} transcription activity could influence the $r^{\mathrm{lac/tet}}$ distribution, as indicated by a series of recent studies *in vitro* and in higher eukaryotic systems [193,198,199].

To examine these possibilities, we first compared the $r^{\mathrm{lac/tet}}$ distribution of the $\lambda\Delta O_L$ strain to that of a control strain $\lambda\Delta O_L P_{RM}^{-} cI^{-}/cI^{trans}$ (Table 3.1, Figure 3.6A and B). In this strain, promoter P_{RM} was mutated to abolish transcription and the *cI* start codon was eliminated, but CI binding to O_R was unaffected (Figure 3.6 C, D and E). In addition, we expressed CI from a plasmid at ~9 times its level in λWT (Table 3.3). We found that the $r^{\mathrm{lac/tet}}$ distributions of the $\lambda\Delta O_L$ and $\lambda\Delta O_L P_{RM}^{-} cI^{-}/cI^{trans}$ strains were indistinguishable (Figure 3.6A and B), demonstrating that the compact $\lambda\Delta O_L$ distribution does not depend on P_{RM} transcription. Furthermore, $r^{\mathrm{lac/tet}}$ distributions for the same $\lambda\Delta O_L P_{RM}^{-} cI^{-}$ strain with or without the CI-expressing plasmid were indistinguishable (Figure 3.6A and B) suggesting that non-specifically bound CI did not interact with specifically bound CI at OR operator sites to condense DNA *in vivo* [193].

Table 3.3 Expression levels of CI in experimental strains.

Strain	N	CI expression level (WLU)
JL5392	3	1.0 ± 0.5
$\lambda\Delta O_L P_{RM}^{-} cI^{-}$	2	0.2 ± 0.1
$\lambda\Delta O_L P_{RM}^{-} cI^{-}/cI^{\mathrm{trans}}$	2	9.3 ± 2.3
$\lambda\mathrm{CI}^{\mathrm{G147D}}$	1	0.2
$\lambda\mathrm{CI}^{\mathrm{G147D}}/cI^{G147D,trans}$	1	11.3

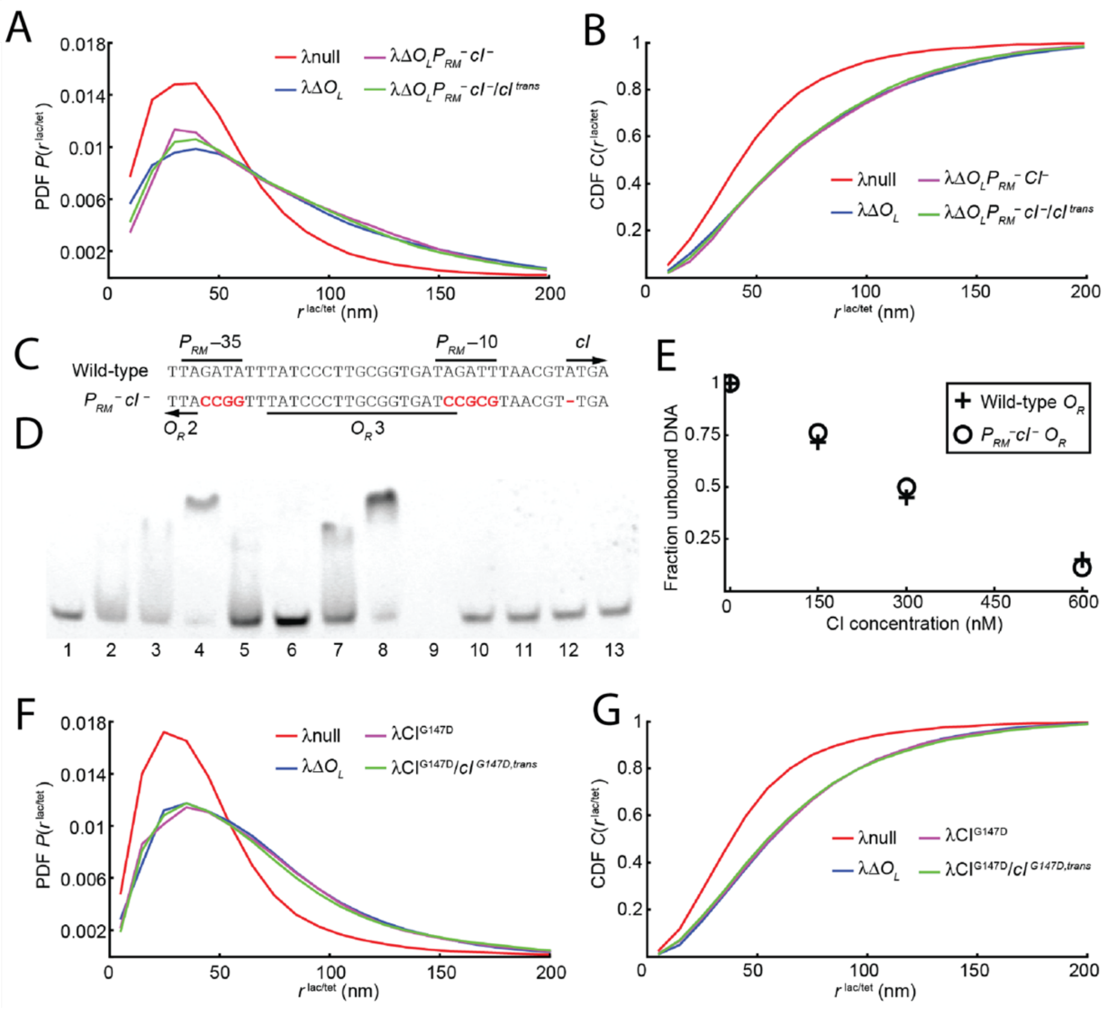


Figure 3.6 Experiments showing the effects of transcription, non-specific CI binding and higher-ordered CI oligomer on DNA looping. (A). End-to-end distance ($r^{\mathrm{lac/tet}}$) distributions (PDF) for λnull (red), λΔO_L(blue), λWT$^{\mathrm{G147D}}$ (black), λΔ$O_L P_{RM}^{-}cI^{-}$ (purple) and λΔ$O_L P_{RM}^{-}cI^{-}/cI^{trans}$ (green). The PDF is estimated for 10-nm bins. (B) Cumulative density of $r^{\mathrm{lac/tet}}$ (CDF) for λnull (red), λΔO_L (blue), λWT$^{\mathrm{G147D}}$ (black), λΔ$O_L P_{RM}^{-}cI^{-}$ (purple) and λΔ$O_L P_{RM}^{-}cI^{-}/cI^{trans}$ (green). The CDF is estimated for 10-nm bins. (C) DNA sequence for the $P_{RM}^{-}cI^{-}$ mutant in comparison to the wild-type sequence. Mutated nucleotides are shown in red. (D) Gel shift assay monitoring the binding of wt CI protein. Lane 1–4: CI at concentrations of 0, 150, 300 and 600 nM binding to a 158-bp DNA fragment (20 nM) amplified from the plasmid pZH107 carrying the wild-type P_{RM} DNA sequence. Lane 5–8: CI at concentrations of 150, 0, 300 and 600 nM (note loading order) binding to a 158-bp DNA fragment (20 nM) amplified from the plasmid pACL007 carrying the $P_{RM}^{-}cI^{-}$ sequence. Lane 9: empty. Lane 10–13: CI at concentrations of 0, 150, 300 and 600 nM binding to a 140-bp DNA fragment (20 nM) amplified from the *E. coli hns* promoter region, which CI does not bind specifically. Reaction mixtures were incubated in a buffer (10mM Tris pH 8.0, 50 mM KCl, 1mM MgCl2, 10% glycerol, 100 ug/ml BSA, 1 mM DTT) at room temperature for 10 minutes. Samples were electrophoresed in Bio-Rad 4–20% Gradient TBE gels (Bio-Rad, Hercules, CA) in a cold room and then stained with Ethidium Bromide for 30 minutes. (E) Fraction of bound DNA (intensity of low-weight band divided by intensity of lane over background) quantified using NIH ImageJ for the gel shown in (D). (F, G) Distributions of $r^{\mathrm{lac/tet}}$ identical in description to those in (A, B) showing strains λnull (red), λΔO_L(blue), λG147D (purple), and λG147D/$cI^{G147D,trans}$ (green).

In vivo observations of DNA looping

We next investigated DNA looping in the context of wild-type and mutant O_R–O_L DNAs. In λWT, the wild-type λ sequence from O_R through O_L was inserted between $lacO^3$ and $tetO^3$. CI could bind all O_R and O_L sites to mediate looping with both octameric and tetrameric CI complexes (Figure 3.1A). In λO_R3^- and λO_L3^-, mutations in O_R3 and O_L3 essentially eliminated CI binding to these operators at lysogenic CI concentrations (Table 3.1) [185,200].

We measured $\vec{\boldsymbol{r}}^{\mathbf{lac/tet}}$ for these three strains and found that $r^{\mathrm{lac/tet}}$ distributions differed significantly from those of the positive and negative controls λnull and λΔO_L ($p < 10^{-3}$, except $p = 0.004$ for λWT and λnull), with $C\left(r^{\mathrm{lac/tet}}\right)$ and $P\left(r^{\mathrm{lac/tet}}\right)$ being intermediate to those of the controls (Figure 3.5 C and D). Mean $r^{\mathrm{lac/tet}}$ values for the three strains also fell in between those of λnull and λΔO_L (Table 3.2). The wild-type strain had lower $\langle r^{\mathrm{lac/tet}}\rangle$ than λO_R3^- and λO_L3^-, and its distribution differed from those of the mutant strains with moderate to high significance (p = 0.001 and 0.048 for λO_R3^- and λO_L3^-, respectively); $r^{\mathrm{lac/tet}}$ distributions for λO_R3^- and λO_L3^- were indistinguishable from each other (p = 0.493). The trend of λnull < λWT < λO_R3^- ≈ λO_L3^- < λΔO_L for $\langle r^{\mathrm{lac/tet}}\rangle$ was reproduced in three independent experiments (Figure 3.2B). Assuming that a DNA molecule in the λWT, λO_R3^- and λO_L3^- strains is in either a looped or unlooped state, the intermediate $\langle r^{\mathrm{lac/tet}}\rangle$ values of the three strains suggested that the fraction of looped DNA molecules (herein termed looping frequency) could be estimated by comparing $r^{\mathrm{lac/tet}}$ distributions of these strains to those of the looped and unlooped controls λnull and λΔ*O*$_L$.

To further investigate whether the observed DNA looping in the λWT, λ*O*$_R$3$^-$ and λ*O*$_L$3$^-$ strains could be abolished by eliminating CI cooperative binding rather than by deleting *O*$_L$, we constructed a control strain λWT$^{\mathrm{G147D}}$ (Table 3.1). This strain differs from λWT by a CI mutation G147D known to be defective in pairwise cooperative interaction [201,202]. Structural evidence suggests that cooperative binding interfaces are shared for pairwise binding to adjacent operator sites and the formation of CI tetramers or octamers via DNA loops [203]. We found that the $r^{\mathrm{lac/tet}}$ distribution of the λWT$^{\mathrm{G147D}}$ strain was indistinguishable from that of λΔO$_L$ (Figure 3.6F and G, Table 3.4), suggesting that DNA looping was eliminated in this mutant CI background. We note that this G147D mutant also diminishes P$_{\mathrm{RM}}$ transcription because of its weakened ability to form a CI tetramer at the O$_R$1 and O$_R$2 sites; hence its expression level is lower than that with wild-type CI (Table 3.3). Therefore, we constructed another control strain (λCI$^{\mathrm{G147D}}$/*cI*^{G147D,trans}), in which the CI$^{\mathrm{G147D}}$ mutant protein was expressed constitutively at ~11 times the CI expression level in λWT from a plasmid transformed into the λCI$^{\mathrm{G147D}}$ strain (Table 3.3). We found that $r^{\mathrm{lac/tet}}$ distribution of this strain was indistinguishable from that of the λΔ*O*$_L$ and the λCI$^{\mathrm{G147D}}$ strains, demonstrating that DNA looping could be abolished by eliminating CI cooperative binding.

Table 3.4 Summary of measurements and samples statistics.

Strain	$r^{\mathrm{lac/tet}}$ measurements	Mean $r^{\mathrm{lac/tet}}$ (nm)	Median $r^{\mathrm{lac/tet}}$ (nm)
λnull	1227	46 ± 1	39 ± 1
$\lambda\Delta O_L$	1201	64 ± 1	58 ± 1
$\lambda\mathrm{CI}^{\mathrm{G147D}}$	873	65 ± 2	58 ± 2
$\lambda\mathrm{CI}^{\mathrm{G147D}}/cI^{G147D,trans}$	1982	64 ± 1	56 ± 1

Estimating DNA looping frequency from $C(r^{\mathrm{lac/tet}})$

To quantitatively examine how operator mutations influence DNA looping, we estimated looping frequencies for λWT, λO_R3^- and λO_L3^- by assuming a simple model. In this model, DNA molecule can only exist in one of two states, looped or unlooped, with $r^{\mathrm{lac/tet}}$ distributions for each state resembling those of the looped and unlooped controls, λnull and $\lambda\Delta O_{L,}$ respectively. Therefore, the distribution $P(r^{\mathrm{lac/tet}})$ or $C(r^{\mathrm{lac/tet}})$ for one of the three strains is the linear combination of that of λnull and $\lambda\Delta O_L$, with their distributions weighted by the looping frequency, f:

$$C(r^{\mathrm{lac/tet}}) = fC(r^{\mathrm{lac/tet}}, \lambda\mathrm{null}) + (1-f)C(r^{\mathrm{lac/tet}}, \lambda\Delta O_L).$$

Using this model, we found that the looping frequency was 79% for λWT, and reduced to 53% for λO_R3^- and 60% for λO_L3^- (results with errors summarized in Table 3.2). The results were indistinguishable within error regardless of whether cumulative or probability density distributions were used, or whether data points from all frames or only the first frame of each molecule's movie were used (Table 3.5). The looping frequencies for λO_R3^- and λO_L3^- were indistinguishable from each other within error, suggesting a similar role of O_R3 and O_L3 in loop formation. Reduced looping frequencies of λO_R3^- and λO_L3^- compared to λWT suggest that while a CI octamer at O_R12 and O_L12 is sufficient to loop DNA, the resulting loop can be further stabilized by an additional CI tetramer only if both O_R3 and O_L3 are intact. These measurements provide the first quantitative *in vivo* estimates of DNA looping frequencies that are independent of gene regulation models.

Table 3.5 Looping frequencies.

Strain	Sample	Distribution	Looping frequency
λWT	All frames	CDF	79 ± 6%
		PDF	78 ± 9%
	First frames	CDF	81 ± 8%
		PDF	76 ± 11%
λO_R3^-	All frames	CDF	53 ± 7%
		PDF	47 ± 9%
	First frames	CDF	56 ± 9%
		PDF	57 ± 10%
λO_L3^-	All frames	CDF	60 ± 7%
		PDF	52 ± 9%
	First frames	CDF	62 ± 8%
		PDF	59 ± 11%

Discussion

In this work, we directly measure the end-to-end separation between two DNA sites separated by only 2.3 kb on the *E. coli* chromosome with unprecedented spatial resolution, and report the first estimates of CI-mediated DNA looping frequencies in live *E. coli* cells. The new method reported here also has the potential to address questions beyond DNA looping, including understanding of chromosome structure and dynamics *in vivo*. In the following, we compare our results with previous work, and discuss unique information provided by our new method.

Differences with *in vitro* looping measurements

Our estimated looping frequencies of 79% for λWT and >50% for λO_R3⁻ and λO_L3⁻ are larger than those observed *in vitro* by TPM and AFM, where looping frequencies at lysogenic CI concentrations were approximately 60% with wild-type operators and 10–40% in the absence of O_R3 and O_L3 [189,191,193]. This likely resulted from differences between naked DNA in an *in vitro* environment and the compact, protein-decorated *E. coli* chromosome in the crowded cellular environment. Factors such as supercoiling and non-specific, "histone-like" DNA-binding proteins (NAPs) could compact DNA and lead to more frequent encounters between O_R and O_L. Our observation that the unlooped λΔO_L DNA was extremely compact (discussed in more detail below) was consistent with this view; this level of compaction (comparable to a polymer with a 3-nm rather than a 50-nm persistence length) could lead to a 50-fold increase in the rate at which O_R and O_L encounter each other [204]. Our looping frequency estimates confirm what were predicted by *in vivo* gene expression experiments—DNA was estimated to loop ~72% of the time for wild-type O_R–O_L DNA and ~69% for DNAs similar to our λO_R3⁻ and λO_L3⁻ constructs [185].

One important assumption we employed in calculating looping frequencies is that that looped and unlooped λWT, λO_R3⁻ and λO_L3⁻ DNA molecules had similar $r^{\mathrm{lac/tet}}$ distributions to those of the looped control λnull and unlooped control λΔO_L, respectively. It is possible that the unlooped states in the λWT, λO_R3⁻ and λO_L3⁻ strains were more compact than that in λΔO_L if, after a DNA loop breaks, O_R–O_L

DNA does not always completely relax before it reforms again. In such a case, looping frequencies estimated using the linear-combination model would be upper limits on the true looping frequencies. Nevertheless, as we show above, our looping frequency estimates broadly agree with expectations from previous studies. Since this simple model only requires one free parameter and gives reasonable results, it is unnecessary to invoke more complicated models.

The short end-to-end separation of $\lambda\Delta O_L$ reflects the high compactness of chromosomal DNA

We observed very small end-to-end separation for the unlooped control ($\langle r^{\text{lac/tet}} \rangle$ = 71 nm). This distance was shorter than expected from modeling the unlooped DNA as a non-interacting worm-chain with an *in vitro* persistence length of 50 nm [205], but consistent with the recently observed extreme bendability of short DNA molecules [206]. A non-interacting chain with an equivalent $\langle r^{\text{lac/tet}} \rangle$ to that of $\lambda\Delta O_L$ would have a persistence length of only 3 nm, which is physically infeasible. Our measurements of indistinguishable conformational distributions in the absence of P_{RM} transcription and the presence of CI overexpression suggest that neither transcription nor non-specifically-bound CI played a major DNA-compacting role in our experiments. Furthermore, *C. crescentus* chromosomal DNA segments of ~5 kb were found to be similarly compact and consistent with Brownian dynamics simulations of supercoiled DNA [74].

We attribute the small end-to-end separation observed for $\lambda\Delta O_L$ to the high compaction of the *E. coli* chromosome in the crowded cellular environment. While the exact molecular mechanisms responsible for compaction remain unclear, previous studies found that *in vitro* binding of the histone-like HU protein [207] and *in vivo* mammalian chromatin packing [208] reduced the apparent persistence length of DNA. Hence, it is possible that nucleoid-associated proteins such as HU may bring distal DNA sites together by protein-protein interactions and/or affect local DNA conformations by introducing bends and relieving torsional strain [209]. Another important factor could be negative supercoiling, which has been shown to compact the chromosomal DNA globally [210]. However, the exact effect of negative supercoiling on a 2.3-kb DNA segment is difficult to predict, because negative supercoiling could also introduce extended, plectonemic structures that promote large separations between DNA sites on relatively short length scales [210].

Potential applications

Our two-color, high-resolution method can be applied to examine how chromosomal location, DNA length, genetic background and growth conditions affect the distance between any two DNA sites on the *E. coli* chromosome. Furthermore, the spatial organization of the *E. coli* chromosome can be determined by systematically measuring $r^{\mathrm{lac/tet}}$ distributions between DNA sites throughout the chromosome. This method is similar to how chromosome conformation capture was used to generate a 3D model of the *C. crescentus* chromosome [8], but with significantly improved spatial resolution and without potential artifacts from fixation.

Materials and Methods

Strain and plasmid construction

The bacterial host strain used for all experiments presented is E. coli K12 MG1655 (Yale Genetic Stock). For more details on strain and plasmid construction see [76].

Cell growth

Bacterial cells were cultured from single colonies from LB plates in M9 minimal media with MEM amino acids (Sigma) and 0.4% glucose, at room temperature (RT) overnight with shaking. Antibiotics were added when appropriate for the strain's resistance (50ug mL $^{-1}$ carbenicillin). Once cells reached mid-log phase (OD600 ~0.3-0.4), cells were centrifuged at RT and resuspended into M9 with 0.4% glycerol and 0.2% L-arabinose and grown for 2hr to induce LacI-mCherry and TetR-EYFP expression. Cells were resuspended in M9 with 0.4% glucose and grown for an additional 2hr before observation to allow for fluorescence protein chromophore maturation.

Sample preparation and imaging

A gel pad made with 3% low-melting-temperature agarose (SeaPlaque, Lonza) in M9 with 0.4% glucose was prepared. Cells were spun-down and resuspended in ~50uL of growth media, ~1uL of cells were immobilized between the gel pad and a coverslip for imaging. An Olympus IX-81 inverted microscope with a 100X oil objective was used, with 1.6x additional amplification. Optosplit II (Andor) was used to split the red and yellow fluorescent channels, images were captured with an Ixon DU-895 (Andor) EM-CCD with a 13-μm pixel width, utilizing the MetaMorph software (Molecular Devices). Illumination at 514 was provided by an argon ion laser (Coherent I-308). Illumination at ~570nm was provided by a rhodamine dye laser (Coherent 599) pumped by 90% of the output from the argon laser. A quarter-wave plate (ThorLabs) was used to circularly polarize excitation light. Emitted light was split using a long-pass filter, and the red and yellow channels were filtered using HQ630/60 and ET540/30 bandpass filters (Chroma).

Measurement of the O_R-O_L separation r $^{lac/tet}$

Images were manually inspected and processed using a custom MATLAB script to identify spots that appeared in both EYFP and mCherry images. Images from multiple strains were displayed in randomized order to avoid bias in manual spot selection. Pixel intensities within 3 pixels of the initial spot location were fitted with a symmetric, 2D Gaussian distribution to estimate spot coordinates. The variance of the fit distribution was constrained to be less than 2 pixels. Spot-fitting error was estimated by scrambling residuals from a fit to the fluorescence data in ten random permutations, adding them to the data, and fitting the resulting images. Reported error for a spot is the standard deviation of the distances between these fits and the initial fit to the raw data. Fitting error distributions are shown in Figure 3.2.

The LacI-mCherry-EYFP tandem dimer in which the two fluorescent proteins were directly fused together was used to acquire fiducial control points to transform between the mCherry and EYFP coordinate systems. A projective transform was calculated from the control points using the cp2tform function in MATLAB. We found that relatively simple, global transformations were sufficient to transform coordinates of fluorescent beads (TetraSpeck, Invitrogen) with ~10-nm registration error in our microscope setup, and did not see any further improvement with a locally weighted transformation used in in vitro two-color experiments [21]. This transformation was also used to generate the overlay images in Figure 3.4. Fluorescent beads were not used as fiducial markers because the beads' emission spectra were different from those of the fluorescent proteins. Analysis was restricted to molecules in which mCherry and transformed EYFP coordinates were separated by less than 200 nm. Separations beyond this threshold were rare (~1% of data, see two-dimensional distributions in Figure S4) and did not correlate with strain identity in any reasonable way. They possibly arose from data in which cells contained two labeled copies of O_R–O_L DNA. After transformation into a uniform coordinate system, $r^{lac/tet}$ was calculated from the mCherry and EYFP coordinates and multiplied by an 81-nm pixel size (resulting from 160 x magnification on a CCD with a 13-mm pixel width). Probability and cumulative distributions $P(r^{lac/tet})$ and $C(r^{lac\text{-}tet})$ were calculated for 10-nm bins using the kernel smoothing probability density estimation (ksdensity) function in MATLAB, restricting the density to positive values and employing a uniform kernel width small enough to follow empirical cumulative density distributions without any systematic errors. Significant differences between $r^{lac/tet}$ distributions were determined using a two-sample Kolmogorov–Smirnov test; two-tailed Student's t tests of sample means returned smaller, more significant p values. Errors in $\langle r^{lac/tet} \rangle$ and <[CI]> were determined by calculating the means of 1,000 bootstrapped samples; the reported error is the standard deviation of the calculated means. Looping frequencies were estimated by least squares fitting of 1,000 bootstrapped distributions (control distributions were also randomized on each iteration) and their error was calculated similarly.

Chapter 4: Visualization of Genomic Sites Using a Reconstituted CRISPR-Cas System - *In Vitro* Characterizations

Introduction

Understanding chromosome organization and dynamics is at the heart of understanding chromosome biology

The makeup of an organism's genome is the blueprint of life. In 2004, the human genome project was successfully completed, and the availability of the linear sequence provided a wealth of possibilities for scientific research; for example, the feat greatly assisted the investigation, identification, and treatment of genomic diseases such as various types of cancer. An important step was made in understanding the complexities of genomic functions through the uncovering of its linear sequence. It soon became clear, however, that the various functionalities of the genome - such as gene expression, DNA replication, and chromosome spatial partitioning - is very complex and could not be inferred from the sequence data alone. For example, many recent studies show that gene regulation is closely tied to the organization of the genome in 3D space on the linear length scale of kilobases to megabases [211,212]. This higher-order spatial organization of the genome is an essential functional component to how genetic material is dynamically maintained and accessed. Due to the need to better understand chromosome organization and dynamics, over the years, various methods have been developed to allow the investigation of both: 1. the spatial positioning of genomic sites inside the cell dimensions, and 2. the spatial relationship between various DNA sites (discussed in the next section). The ultimate goal is to elucidate the spatial organization of DNA, and understand the functional correlation between chromosome architecture and gene regulation.

DNA spatial organization can be probed via imaging or 3C based techniques

There are currently a number of ways to probe DNA conformation in eukaryotic cells in a sequence specific way (Table 4.1). Fluorescent imaging techniques have been a particularly valuable mainstay method for labeling specific DNA sites. This type of labeling has the advantage of allowing visualization of DNA positions in single cells, and often in an intact cell, which gives additional information about relative positioning of DNA sites with respect to the nuclear boundary. Many of the classical techniques based on dye-labeled oligos require prior fixation of the cells, which can potentially introduce fixation artifacts, and mask relevant interactions that have a lower frequency of occurrence (in a population of cells, or in a single cell over time)[34,35,38]. Alternatively, a site of interest can be labeled using a tandem array of operator sequences (Fluorescent Reporter Operator System (FROS)) [213]. This method is a low throughput way of imaging gene sites, but has the advantage of being a live-cell compatible technique. Chromosomal Conformation Capture (3C), and its many derivative methods, represent a third approach, and depend on chemical fixation to capture the 3D spatial organization of the chromosome; further digestion and ligation reveal interactions between linearly distal sites that can be read-out by PCR or sequencing [28,61,214]. 3C-based techniques look at genome interactions in a bulk sample of cells and thus suffer from the loss of single-cell level information, which is especially relevant in an unsynchronized population of cells. Also, 3C and its derivative techniques inherently lack positioning information relative to the physical space of the nucleus.

Table 4.1 Review of current techniques for sequence specific investigation of 3D genome conformation in eukaryotes. *Last row (gray), method currently under development in this study.*

Method	Live-cell	Principle	Strength	limitations	Ref.
3C and derivatives	No	Fixation and cross-linking of genome interactions, followed by PCR/sequencing	High throughput, one pot reaction that detects all long-range interactions (>10kb)	Bulk cell study; inference of interaction via cross-linking	61
Endogenous binding protein	Yes	Visualizing DNA sequences with endogenous binding protein by tagging these proteins	Easy to implement	Limited sequences with endogenous binding proteins	60
smFISH	No	Multiple dye-labeled oligo probes binding to complementary targeting site sequence	Applicable to probing domain structures	Requires fixation, expensive, hybridization efficiency is variable	34,35,215
FROS ParB/INT	Yes	Detecting of tandem operator sites via the binding of fluorescent protein fused to DNA-binding proteins	Live-cell compatible, DNA tracking, strong signal:noise, orthogonal systems available	Needs insertion of artificial binding sites, difficult to map many sites in the same experiment	213
TALE-FP ZFP-FP	Yes	Protein is made up of modular domains engineered to recognize specific DNA sequences	Can be designed based on 'coded' modular domains	Only practical for labeling repetitive DNA sequences, extensive screening needed for ZFPs	216-219
CRISPR: dCas9-FP	Yes	Fluorescently labeled dCas9 protein is used to bind specific genomic sites via the targeting ability of sgRNAs	Live-cell compatible, DNA tracking, affordable, many orthogonal CRISPR systems	Targeting sequence is restricted by PAM, binding of dCas9 could potentially perturb functionality of DNA, limited to single color for single CRISPR system	220-223
CRISPR: aptamer modified sgRNA	Yes	sgRNAs are modified with insertions of hairpin aptamers that could be bound with binding protein-FP fusions	Live-cell compatible, DNA tracking, multiplexing within a single CRISPR system (simultaneous detection of multiple sites)	Need to integrate combinatorial system into cells to express a variety of sequence modified sgRNAs and binding proteins	224,225

CRISPR: Dye-labeled RNA oligos	Yes	Synthesized RNA oligos are dye-labeled and used to target dCas9 to binding to specific DNA sites	Simple in vitro delivered system that can be used in wt cells, easily changed to probe different DNA sites	Efficiency of detection is variable depending on target site, multiplexing is ultimately limited by dye choices	This project

CRISPR-Cas9 allows for the targeting of sequence specific DNA sites via guide RNA

The discovery and characterization of CRISPR systems and the rapid revolution in adapting CRISPR as a genome editing tool reflects the unprecedented ease and affordability of using this RNA-guided system compared to prior alternatives such as zinc-finger nucleases (ZFNs), and transcription activator-like effector nucleases (TALENs) [226]. The relative fidelity of Cas9 in DNA target recognition and the ability to target almost any genomic site (as long as there is a PAM sequence nearby, NGG for *S. pyogene*), makes Cas9 an ideal tool for the sequence-specific targeting of DNA [227,228]. The Cas9 protein can be engineered to lose its cleavage function, allowing the resulting deactivated Cas9 (dCas9) to act as an RNA-guided, sequence-specific DNA binding protein (Figure 4.1) [228,229]. It has been shown that dCas9 alone can be used to physically block RNAP, resulting in the inhibition of transcription [230]. Furthermore, when fused to a wide variety of partners, including repressors, activators, histone demethylases, and histone acetyltransferases, dCas9 fusions can help bring different functional domains within close proximity of the DNA locus of interest, thereby allowing for targeted transcription and epigenetic regulation [230,231].

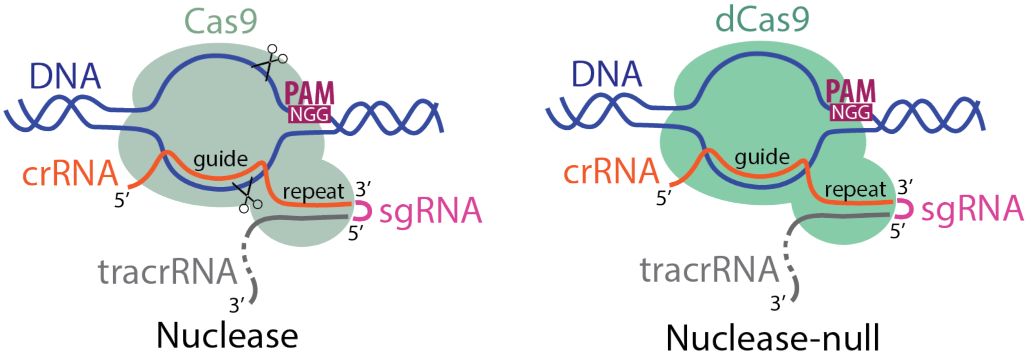


Figure 4.1 Schematic of Cas9 vs. dCas9 recognition of DNA guided by RNA. Left: the entire CRISPR-Cas9 complex consists of the protein Cas9, and either two short RNAs: crRNA and tracrRNA, or a single sgRNA. The crRNA contains a variable guide region that is complementary to the targeting region in DNA, and tracrRNA acts mostly as a structural component and its sequence is universal for a single species of Cas9 protein. The single piece of sgRNA (single-guide RNA) is an experimentally modified RNA which connects the crRNA and tracrRNA via a newly introduced tetra-loop, for ease of endogenous expression. Scissors indicate where double stranded cleavage occurs via two separate nuclease domains. Right: dCas9 has been engineered to have no nuclease activity.

CRISPR-Cas9 system has been adapted for visualization of DNA loci

The CRISPR-Cas9 system from *s. pyogene* has also recently been used to visualize chromosomal sites in mammalian cells [232]. Here, the cleavage disabled dCas9 protein is fused with a fluorescent protein, and is directed to bind to specific DNA sequences using targeting sgRNAs (single-guide RNAs) [220] (Figure 4.2A). Single-guide RNA is a fusion construct that combines the DNA targeting crRNA and the universal tracrRNA; the crRNA contains the variable sequence that hybridizes with DNA targets, while the tracrRNA serves as a structural component of the protein-RNA complex [227,233,234]. The CRISPR system can be easily implemented for visualization of specific DNA sites, due to the ease of programming dCas9 to target different genomic sites through the endogenous expression of both dCas9 and sgRNAs. In comparison, FROS techniques would require the insertion of long artificial binding sites into the genome, and smFISH techniques would require cell fixation, and the purchase of expensive target specific probes. dCas9-FP fusions provide an unprecedented way to visualize DNA sites in a living cell without a priori sequence alteration to the target site.

However, visualization based on dCas9-FP fusions is inherently limited to visualizing one DNA site at a time. To allow for the detection of multiple genomic sites simultaneously, there have been two general directions of advancement. These include: 1. The use of orthogonal CRISPR systems (Figure 4.2A) and 2. The use of modified sgRNAs that contain multiple RNA aptamers (Figure 4.2B, Table 4.1). Orthogonal CRISPR-Cas9 systems from other bacterial species that have distinctive PAMs and sgRNAs have been experimentally shown to be compatible for multi-color imaging of up to three DNA sites [222,223] (Figure 4.2A). To further expand the multiplexing capabilities of the CRISPR imaging system, RNA aptamer insertion on the sgRNA has allowed for the simultaneous detection of up to seven sites [225] (Figure 4.2B). The disadvantage of both strategies lies in the need to have the cell express a large number of protein and RNA components at optimal levels, which makes the experimental design and implementation more time consuming and difficult to fine-tune [232]. Additionally, since both RNA and protein components of the CRISPR-Cas9 system are to be endogenously expressed in the cell, there is a realistic upper limit to the number of components that can be further added to simultaneously detect even more DNA sites.

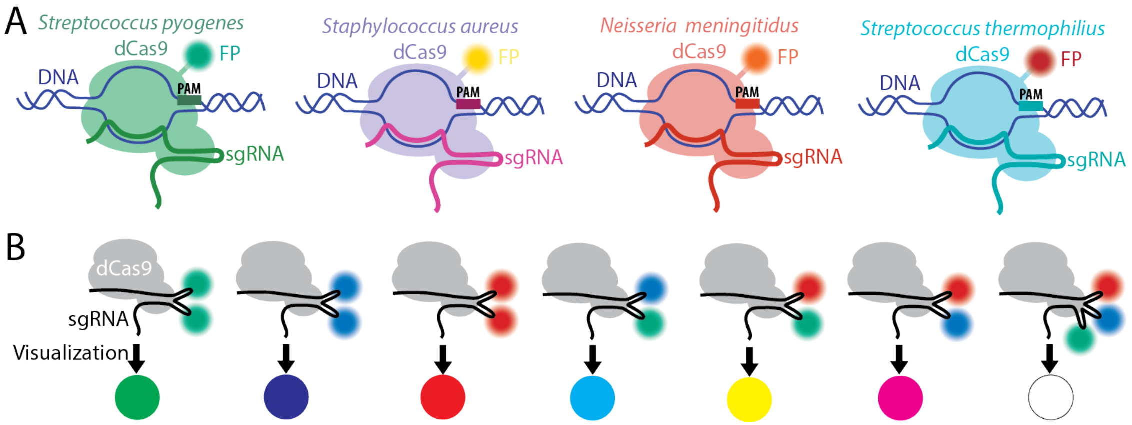


Figure 4.2 Current CRISPR-Cas9 based methods for simultaneous detection of multiple DNA sites. (A) Multiplexing can be achieved using orthogonal dCas9 from multiple bacterial species with distinctive sgRNA and PAM requirements, visualization is through fluorescent protein fusions to dCas9. (B) Different RNA aptamers are incorporated into the sgRNA scaffold in various paired configurations, which allows for the binding of different sets of fluorescent protein fusions. The identity of the sgRNA and there by the identity of the targeted DNA site is uncovered through the read-out of the combinatorial colors.

A reconstituted *in vitro* delivery system would improve upon current methods

All current methods allowing for the CRISPR-Cas9 based labeling of multiple DNA sites in live cells requires the balanced endogenous expression of a large number components. In addition, all of these techniques also depend on some type of fluorescent protein fusion for visualization (dCas9-FP, or RNA aptamer binding protein-FP, Table 4.1), which inherently has a lower signal-to-noise ratio compared to dye-based labeling systems, and potentially higher background due to variability in expression levels and relatively long protein lifetimes [232]. We aim to develop a CRISPR-Cas based DNA labeling system that would require minimal cell line preparation, and that can be quickly implemented in wildtype cells. This method would retain the advantage of being applicable in live cells and can be multiplexed for probing multiple gene sites in an unprecedented high-throughput manner (Table 4.1).

We propose to develop the CRISPR-Cas DNA labeling protocol using dye-labeled RNA components; the entire reconstituted complex with dye-labeled RNA oligos would be introduced into the cell nucleus by *in vitro* delivery (Figure 4.3). The RNA oligo component could be a single species of dye-labeled sgRNA, this would be ideal for targeting repeat sequences on the chromosome such as the telomere region [220] (Figure 4.3A). For targeting a sequentially variable region, a number of unlabeled crRNA species with various targeting sequences could be used by pairing with a universal dye-labeled tracrRNA oligo (Figure 4.3B). This two-piece (2pc) tracrRNA-crRNA setup would greatly reduce the cost of targeting sequentially variable regions on the genome, compared to using a number of longer variable sgRNAs (Figure 4.3B). Alternatively, the CRISPR-Cas complex could be used to target multiple genomic sites with variable sequences, by using spectrally separate dyes to label the variable crRNA and providing an unlabeled tracrRNA component. The sequential color readout from this combinatorial labeling approach would reflect the sequence identity of the underlying DNA at the same time as it reports on the spatial positioning of this genomic site (Figure 4.3C). Traditional labeling approaches are capable of detecting a limited number of molecular species simultaneously, due to the spectral overlap between fluorophores, which in turn limits the number of colors that can be practically used.

This method of combinatorial labeling combined with super resolution imaging greatly increases the capacity for multiplexing in single cells [235], and has already been successfully implemented to spectrally 'bar-code' mRNA using dye-labeled FISH probes [236].

In this chapter, we will detail our efforts in characterizing and optimizing the reconstituted CRISPR-Cas9 system first using *in vitro* assays, as a proof of principle.

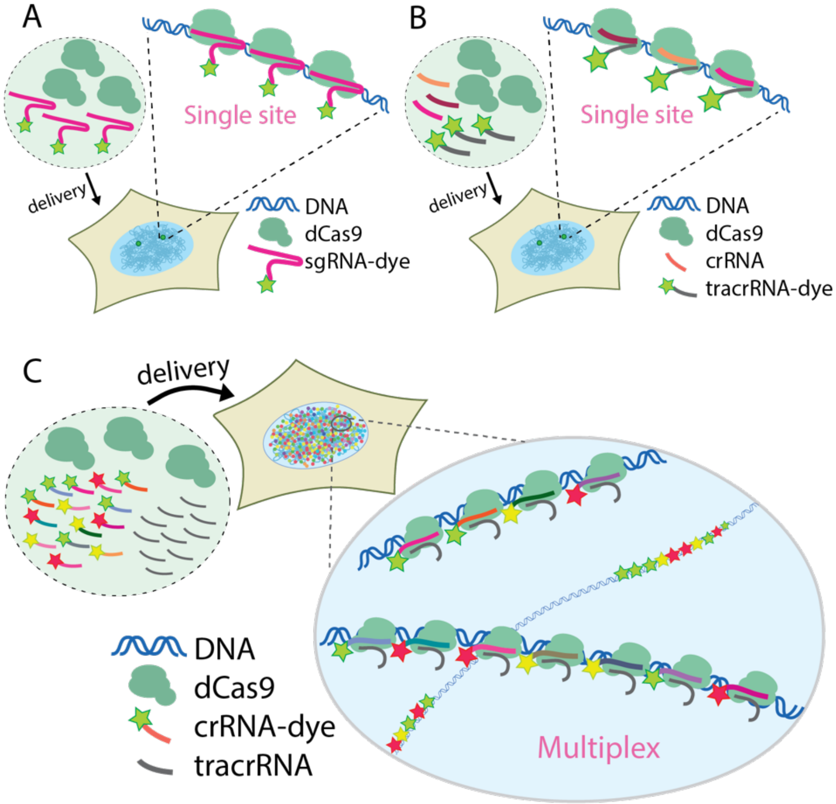


Figure 4.3 Schematic of *in vitro* reconstituted CRISPR-Cas9 DNA labeling system. (A). A delivery system consisting of dCas9 and dye-labeled sgRNA; this setup is optimal for targeting repeat sequences, where only a single targeting sgRNA is needed. (B). A delivery system consisting of dCas9, a number of variable crRNA, and a single universal dye-labeled tracrRNA; this setup is optimal for targeting a single DNA site target with variable sequence. (C). A delivery system with dCas9, variable dye-labeled crRNA, and universal unlabeled tracrRNA, this setup is for combinatorial labeling of a large number of target DNA sequences.

Results

Design of RNA oligos for DNA targeting using the CRISPR-Cas9 complex

First, we set out to conduct proof-of-principle experiments, where we aim to use our dye-labeled CRISPR-Cas9 system to recognize sites previously probed in the literature. Previously, an endogenous fluorescent protein fusion of dCas9 and endogenously expressed sgRNAs were used to visualize long genomic repeat sites such as telomeres and MUC4 [220]. We designed a number of different RNA oligos that we could use to test the functionality of the reconstituted CRISPR-Cas9 complex (see the list of RNA oligos in Table 4.2). Ultimately, for application in live cells, we are only interested in observing dCas9 binding to DNA in a site directed manner; however, we first set out to test our *in vitro* reconstituted system using Cas9 by *in vitro* cleavage assays. Although Cas9 binding and cleavage are not completely correlated [237], we can use cleavage efficiency of the substrate DNA as a gauge for whether the functionality of our system is comparable to the previously characterized sgRNA system.

Table 4.2 RNA oligo designs for Cas9/dCas9. *(*) phosphorothioate modified backbone; (m) 2'-Omethyl bases. In white, RNAs are purchased from Dharmacon. In gray shading (oligos A-F), RNAs were made from in vitro transcription reactions via T7 RNA polymerase. Shorthand names are used in our lab for simplicity.*

Name	**Description**	**Sequence**
1	**tracrRNA**	5' CAG CAU AGC AAG UUA AAA UAA GGC UAG UCC GUU AUC AAC UUG AAA AAG UGG CAC CGA GUC GGU GCU UUU UUU 3'
2	**5′ DY547 tracrRNA**	5' (DY547) CAG CAU AGC AAG UUA AAA UAA GGC UAG UCC GUU AUC AAC UUG AAA AAG UGG CAC CGA GUC GGU GCU UUU UUU 3'
3	**3′ DY547 tracrRNA**	5' CAG CAU AGC AAG UUA AAA UAA GGC UAG UCC GUU AUC AAC UUG AAA AAG UGG CAC CGA GUC GGU GCU UUU UUU (DY547) 3'

4	**GFP crRNA**	5' CGU GCU GAA GUC AAG UUU GAG UUU UAG AGC UAU GCU G 3'
5	**Telomere crRNA**	5' GUU AGG GUU AGG GUU AGG GUU AGU UUU AGA GCU AUG CUG 3'
6	**5′ Cy3 modified telomere crRNA**	5' (Cy3) mG(*)mU(*)mU(*) AGG GUU AGG GUU AGG GUU AGU UUU AGA GCU AUG (*)mC(*)mU(*)mG 3'
7	**Modified tracrRNA**	**5'** mC(*)mA(*)mG(*) CAU AGC AAG UUA AAA UAA GGC UAG UCC GUU AUC AAC UUG AAA AAG UGG CAC CGA GUC GGU GCmU(*)mU(*)mU(*)U **3'**
8	**5′ Cy3 telomere crRNA**	**5'** (Cy3) GUU AGG GUU AGG GUU AGG GUU AGU UUU AGA GCU AUG CUG **3'**
9	**5′ Cy3 modified MUC4-E3 crRNA**	**5′** (Cy3) mG(*)mU(*)mG(*) GCG UGA CCU GUG GAU GCU GGU UUU AGA GCU AUG (*)mC(*)mU(*)mG 3'
10	**Modified telomere crRNA**	5′ mG(*)mU(*)mU(*) AGG GUU AGG GUU AGG GUU AGU UUU AGA GCU AUG (*)mC(*)mU(*)mG 3′
A	**sgMUC4-E3**	**5′** GGU GGC GUG ACC UGU GGA UGC UGG UUU AAG AGC UAU GCU GGA AAC AGC AUA GCA AGU UUA AAU AAG GCU AGU CCG UUA UCA ACU UGA AAA AGU GGC ACC GAG UCG GUG CUU UUU UU **3′**
B	**3′ Cy3 sgMUC4-E3**	**5′** GGU GGC GUG ACC UGU GGA UGC UGG UUU AAG AGC UAU GCU GGA AAC AGC AUA GCA AGU UUA AAU AAG GCU AGU CCG UUA UCA ACU UGA AAA AGU GGC ACC GAG UCG GUG CUU UUU UU (Cy3) **3′**
C	**sgTelomere**	**5′** GGU UAG GGU UAG GGU UAG GGU UAG UUU AAG AGC UAU GCU GGA AAC AGC AUA GCA AGU UUA AAU AAG GCU AGU CCG UUA UCA ACU UGA AAA AGU GGC ACC GAG UCG GUG CUU UUU UU **3′**
D	**3′ Cy3 sgTelomere**	**5′** GGU UAG GGU UAG GGU UAG GGU UAG UUU AAG AGC UAU GCU GGA AAC AGC AUA GCA AGU UUA AAU AAG GCU AGU CCG UUA UCA ACU UGA AAA AGU GGC ACC GAG UCG GUG CUU UUU UU (Cy3) **3′**
E	**GFP crRNA**	**5′** GGC GUG CUG AAG UCA AGU UUG AGU UUU AGA GCU AUG CUG **3′**
F	**Telomere crRNA**	**5′** GGU UAG GGU UAG GGU UAG GGU UAG UUU UAG AGC UAU GCU G **3′**

The sequence design of Telomere and MUC4-E3 sgRNAs was based on the endogenous expression system from Bo Huang's group [220]. Both target highly repetitive sequences in the mammalian genome and serve as a starting point for our method development. In the design of our 2-pc crRNA and tracrRNA system, we eliminated the tetra-loop from the sgRNA design (Figure 4.1), and retained the extended stem region from previously-reported optimizations [220]. Since we set out to test the effectiveness of the dye-labeled 2-pc system in specific DNA sequence recognition, we included RNA oligos with differentially positioned dye labels (5' or

3'), we tested for dye-labeling on either the crRNA or the tracrRNA, and we also included sgRNA controls that target the same DNA sequence as a reference point. Additionally, to mitigate against the possible degradation of RNA oligos during delivery procedure into live cells (for the future application of our method), we designed RNA oligos with chemical modifications on the 3' and 5' termini similar to those previously described [238]; specifically, the last three nucleotides on both ends of the oligo have chemical modifications consisting of 2'-Omethyl 3'phosphorothioate moieties (MS) (Table 4.2). It was shown that chemically modified sgRNAs significantly improved the genome editing efficiencies of Cas9 in comparison to endogenously expressed sgRNAs [238], likely due to their resistance to degradation inside the cell. Importantly, it was shown that the MS modified sgRNAs are non-toxic and could be delivered into human cell lines by nucleofection. Again, we can use cleavage efficiency of target DNA as a read-out for whether RNA-protein complex formation and DNA sequence recognition is compromised by the use of dye-labeled and chemically modified RNA oligos in our 2-pc tracrRNA-crRNA system.

In vitro DNA cleavage experiments with RNA oligos and Cas9

Both unlabeled and dye-labeled sgMUC4 and sgTelomere is specific in targeting DNA for cleavage via Cas9

First, we performed an *in vitro* cleavage assay using previously implemented 1-pc sgRNAs. We tested the DNA targeting and cleavage capability of the telomere-targeting sgRNA (sgTelo) and MUC4-targeting sgRNA (sgMUC4), both synthesized by T7 *in vitro* transcription (see materials and methods for details, Table 4.2). To test whether dye-labeling of the sgRNAs had any adverse effects on DNA target recognition and cleavage, we also created 3' Cy3-labeled sgTelo and 3' Cy3-labeled sgMUC4 using an enzymatic labeling approach (see materials and methods for details). The DNA substrate for cleavage was a linearized plasmid with the targeting sequence adjacent to the appropriate PAM motif (NGG), unless otherwise stated. To test the specificity of DNA sequence targeting, we also included a control with a DNA substrate that has the MUC4 targeting sequence, but without the adjacent PAM motif. We saw that both the unlabeled and 3' Cy3-labeled sgMUC4 guided Cas9 cleaved the MUC4 DNA substrate almost to completion, whereas the MUC4 DNA substrate without the PAM motif displayed no cleavage, showing that nuclease activity is PAM motif dependent and site specific (Figure 4.4A).

Similarly, when we tested the DNA targeting capabilities of the sgTelo RNA, we found that with or without 3' Cy3 labeling, sgTelo was capable of directing Cas9 to cleave telomere sequence containing DNA substrate (with PAM) specifically, and fairly efficiently (Figure 4.4B). Also, when compared with sgMUC4, the telomere targeting sgTelo was able to achieve a lesser degree of substrate cleavage under the same experimental conditions (see materials and methods), showing that some variations in DNA sequence recognition, binding, and cleavage efficiency were likely due to the particular sequence of the target site. Overall, while native telomere targets are a suitable choice for dCas9 directed visualization of DNA *in vivo* due to its repeated nature, the single copy telomere sequence itself was not an ideal cleavage substrate for Cas9, in comparison to MUC4.

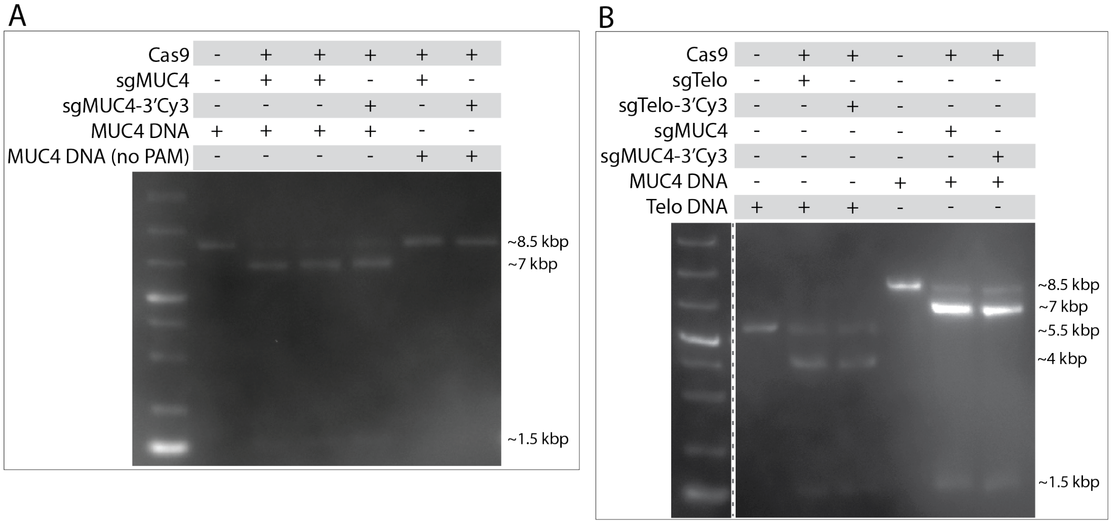


Figure 4.4 *In vitro* cleavage of DNA with Cas9 and sgRNAs. (A) Cleavage assay gel showing reactions of MUC4 DNA cleavage with Cas9 complexed with both unlabeled and dye-labeled sgMUC4. (B). Cleavage assay gel showing reactions of telomere (Telo) and MUC4 DNA cleavage with Cas9 complexed with both unlabeled and dye-labeled sgTelo and sgMUC4.

2-pc system (crRNA with tracrRNA) is capable of specifically targeting DNA for cleavage via Cas9

For the eventual development of our reconstituted DNA labeling system into a multi-site method, it is imperative that we streamline RNA oligo design and synthesis and lower production costs. The universal nature of the tracrRNA in the 2-pc system allows us to just synthesize the shorter variable crRNA for the purpose of targeting multiple gene sites (Figure 4.3B and C). We tested the ability of the 2-pc tracrRNA-crRNA system to direct Cas9 to cleavage of specific sequences. The cleavage assay showed that a GFP-sequence targeting tracrRNA-crRNA pair successfully directed Cas9 to cleave the GFP DNA substrate almost to completion. Both T7 *in vitro* transcribed GFP crRNA and commercially synthesized GFP crRNA showed similar activity (Figure 4.5A). We also saw that the telomere sequence targeting tracrRNA-crRNA pair was able to direct Cas9 to cleave the telomere DNA target. This cleavage activity was also site-specific, since the telomere DNA substrate was not cleaved in the reaction with the GFP sequence targeting tracrRNA-crRNA pair (Figure 4.5B).

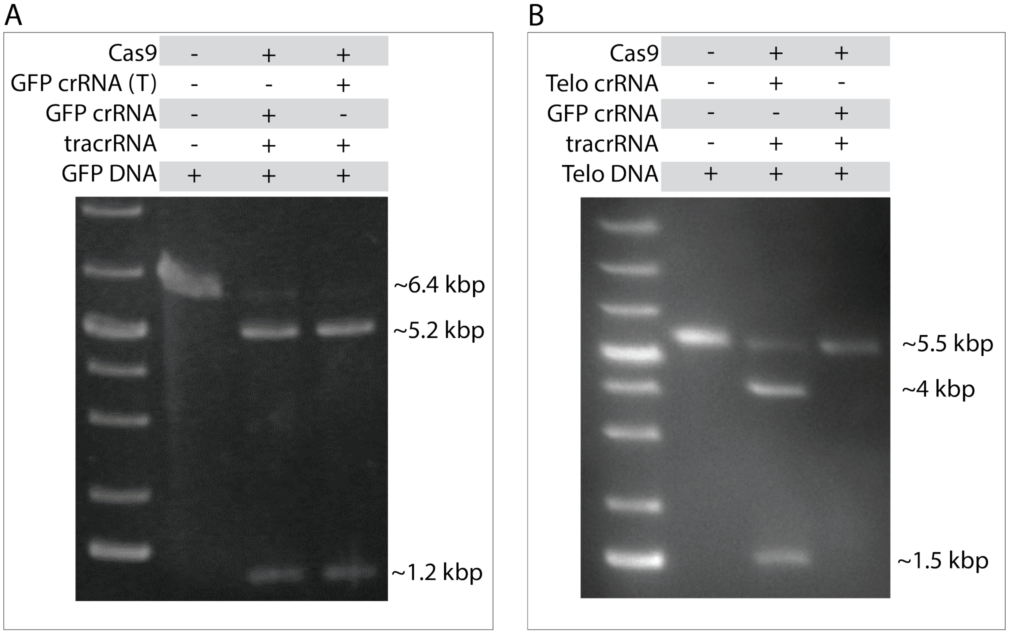
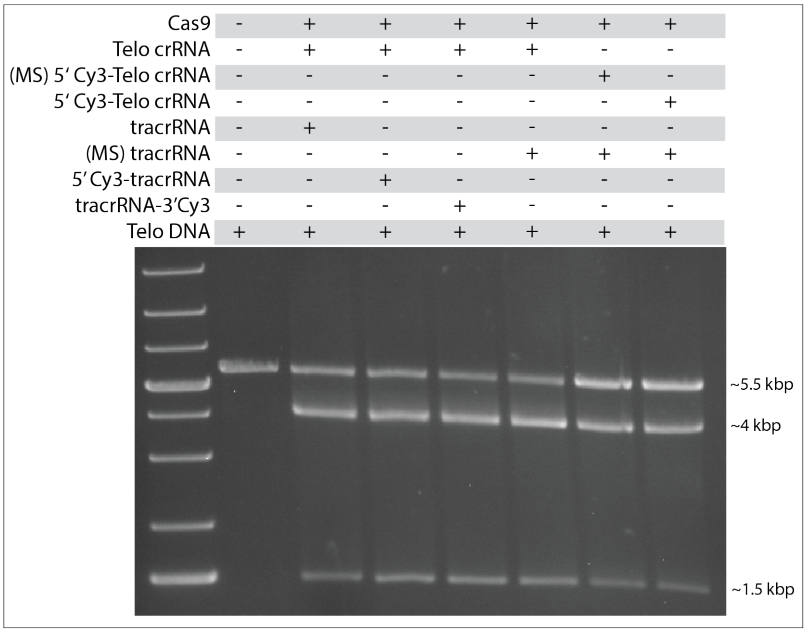


Figure 4.5 *In vitro* cleavage of DNA with Cas9 and tracrRNA-crRNA targeting telomere and MUC4 sequences. (A) Cleavage assay gel showing reactions of GFP DNA cleavage with Cas9 complexed with both commercially synthesized GFP crRNA and transcribed GFP crRNA, paired together with tracrRNA. (B). Cleavage assay gel showing reactions of telomere (Telo) DNA cleavage with Cas9 complexed with tracrRNA and crRNA targeting the telomere sequence, or GFP sequence.

Dye-labels and chemical modifications on RNA oligos have minimal adverse effects on cleavage efficiency of telomere target sequence using the 2-pc system

Dye-labels need be placed on tracrRNA to visualize repeat sequences, or alternatively, crRNA could be labeled with multiple dye colors for multiplexing the detection of multiple genomic sites (Figure 4.3). We set out to test a variety of dye-labeling sites, for example, at the 5' end of crRNA, and at the 5' or 3' ends of tracrRNA. As previously described, chemically modified RNA oligos are better substrates for *in vitro* reconstituted RNA-protein complexes, since they mitigate degradation of RNA oligos during the experimental procedure and after complex delivery into cells. We also tested whether the chemical modifications introduced at the ends of the oligos bring any changes to Cas9 cleavage of substrate DNA. Telomere DNA substrate targeting and cleavage with Cas9 using the dye-labeled and chemically modified tracrRNA-crRNA showed that in comparison to unlabeled and unmodified tracrRNA-crRNA, the biggest drop in cleavage efficiency (~5% calculated from measurement of DNA band intensities) was from using (MS) tracrRNA - (MS) 5' Cy3-Telo crRNA, and (MS) tracrRNA - 5' Cy3-Telo crRNA pairs (Figure 4.6, see Table 4.3). It seems that some perturbation was induced by the 5' dye-labels on Telo crRNA, although most of the cleavage activity was retained.



Cas9	-	+	+	+	+	+	+
Telo crRNA	-	+	+	+	+	-	-
(MS) 5′ Cy3-Telo crRNA	-	-	-	-	-	+	-
5′ Cy3-Telo crRNA	-	-	-	-	-	-	+
tracrRNA	-	+	-	-	-	-	-
(MS) tracrRNA	-	-	-	-	+	+	+
5′Cy3-tracrRNA	-	-	+	-	-	-	-
tracrRNA-3′Cy3	-	-	-	+	-	-	-
Telo DNA	+	+	+	+	+	+	+

Figure 4.6 *In vitro* cleavage of DNA with Cas9 and tracrRNA-crRNA with dye-labels and chemical modifications targeting telomere sequence. Cleavage assay gel showing reactions of Telo DNA cleavage with Cas9 complexed with a variety of tracrRNA-crRNA pairs, some containing dye-labels or MS modifications.

Table 4.3 Cleavage efficiencies measured for 2-pc system targeting telomere DNA, with various combinations of tracrRNA and crRNA types.

DNA	**Telo**	**Telo**	**Telo**	**Telo**	**Telo**	**Telo**
tracrRNA	tracrRNA	5’ Cy3-tracrRNA	tracrRNA-3’Cy3	(MS) tracrRNA	(MS) tracrRNA	(MS) tracrRNA
crRNA	Telo crRNA	Telo crRNA	Telo crRNA	Telo crRNA	(MS) 5’Cy3-Telo crRNA	5’Cy3-Telo crRNA
Cleavage efficiency	66.7%	68.3%	69.2%	68.8%	62.9%	61.6%

2-pc tracrRNA-crRNA system is better than 1-pc sgRNA in targeting Cas9 for telomere DNA cleavage

Under the same reaction conditions, interestingly, the 2-pc tracrRNA-crRNA system was better at directing Cas9 for telomere DNA substrate cleavage compared to the 1-pc sgRNA. We saw that overall the 2-pc system has higher cleavage efficiency, and that even the dye-labeled and MS-modified 2-pc system lead to more cleaved telomere DNA substrate compared to sgTelo RNA (unlabeled and Cy3 labeled, Figure 4.7). This result could be a general characteristic related to the nature of the 2-pc vs 1-pc set-up, or another possibility is that this trend is specific to the telomere targeting sequence and not generally applicable to other sequences.

Cas9	-	+	+	+	+	+	+
Telo crRNA	-	+	-	-	-	-	-
(MS) 5′Cy3-Telo crRNA	-	-	-	-	+	+	+
tracrRNA	-	+	-	-	+	-	-
5′ Dy547-tracrRNA	-	-	-	-	-	+	-
tracrRNA-3′Dy547	-	-	-	-	-	-	+
sgTelo	-	-	+	-	-	-	-
sgTelo-3′Cy3	-	-	-	+	-	-	-
Telo DNA	+	+	+	+	+	+	+

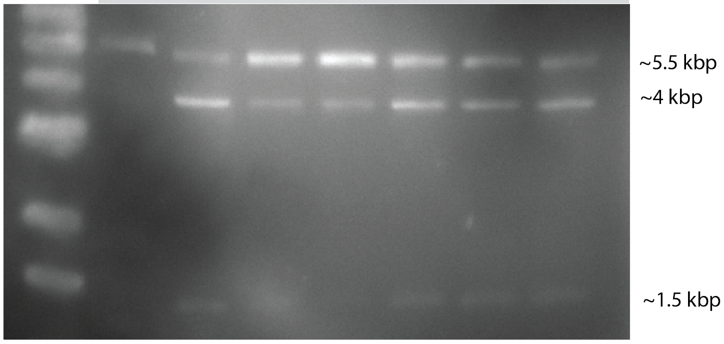


Figure 4.7 *In vitro* cleavage of Telo DNA with Cas9 and either the Telo sgRNA or telomere targeting tracrRNA-crRNA. Cleavage assay gel showing reactions of Telo DNA cleavage with Cas9 complexed with a variety of tracrRNA-crRNA pairs, some containing dye-labels or MS modifications, and in addition, 1-pc sgRNA with or without Cy3 labels has been used as a comparison.

Dye-labels and chemical modifications on RNA oligos have small effects on the cleavage efficiency of MUC4 target sequence using the 2-pc system

Similar to previously described, we tested the effect of dye-labels and chemical modifications on the tracrRNA-crRNA 2-pc system that target Cas9 to the MUC4 substrate DNA. We saw that chemical modifications or dye-labels caused minimal changes (~1%) in cleavage efficiency of the MUC4 DNA substrate, among the different pairs of tracrRNA and crRNA tested (Figure 4.8, see Table 4.4).

Cas9	-	+	+	+	+	+
(MS) 5′ Cy3-MUC4 crRNA	-	+	+	+	+	+
tracrRNA	-	+	+	-	-	-
(MS) tracrRNA	-	-	-	-	-	+
5′ Cy3-tracrRNA	-	-	-	+	-	-
tracrRNA-3′Cy3	-	-	-	-	+	-
MUC4 DNA	+	-	+	+	+	+
MUC4 DNA (no PAM)	-	+	-	-	-	-

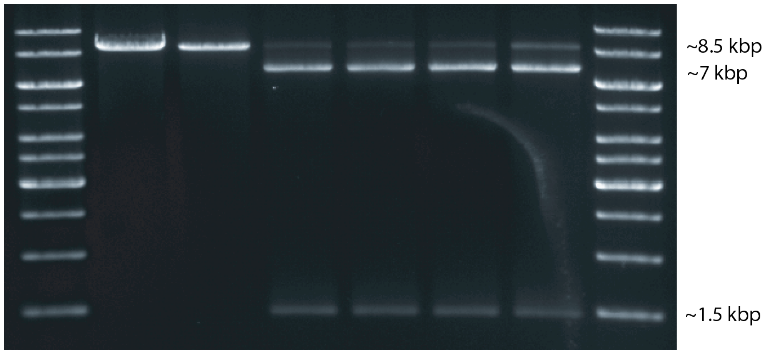


Figure 4.8 *In vitro* cleavage of MUC4 DNA with Cas9 and tracrRNA-crRNA containing dye-labels and chemical modifications. Cleavage assay gel showing reactions of MUC4 DNA cleavage with Cas9 complexed with a variety of tracrRNA-crRNA pairs, some containing dye-labels or MS modifications.

Table 4.4 Cleavage efficiencies measured for 2-pc system targeting MUC4 DNA, with variant combinations of tracrRNA and crRNA types.

DNA	**MUC4 (no PAM)**	**MUC4**	**MUC4**	**MUC4**	**MUC4**
tracrRNA	tracrRNA	tracrRNA	5’Cy3-tracrRNA	tracrRNA-3’Cy3	(MS) tracrRNA
crRNA	(MS) 5’Cy3-MUC4 crRNA	(MS) 5’Cy3-MUC4 crRNA	(MS) 5’Cy3-MUC4 crRNA	(MS) 5’Cy3-MUC4 crRNA	(MS) 5’Cy3-MUC4 crRNA
Cleavage efficiency	0%	88.7%	88.8%	89.2%	88.2%

In vitro DNA cleavage experiments with CRISPR-Cpf1

Dye-labels and chemical modifications have small effects on cleavage efficiency of CRISPR-Cpf1

Lastly, we investigated the ability of an analogous RNA guided CRISPR-Cas system - CRISPR-Cpf1 - to cleave target DNA using dye-labeled and chemically modified RNA oligos. Importantly, Cpf1 uses just a single, shorter RNA for sequence specific DNA targeting, and the Cpf1 protein is a smaller endonuclease compared to Cas9 (Figure 4.9A) [239-241]. Overall, CRISPR-Cpf1 is a simpler system compared to CRISPR-Cas9, and could potentially prove to be another system we can utilize for development of our CRISPR based DNA visualization methodology.

We tested cleavage activity of Cfp1 from three different species: As (*Acidaminococcus sp.*), Lb (*Lachnospiraceae*) and Fn (*Francisella novicida*). All three species variants share the similar PAM motif requirements (TTN), and need around 17-19 nt of variable sequence for optimal DNA hybridization and targeting [239]. For Cpf1, we designed two targeting sequences for the α-satellite repeats in the mammalian genome, α-1 and α-2 [222] (see Table 4.5); both RNA oligos were dye-labeled with Cy3 at the 3' end and chemically modified. We saw that overall, cleavage was close to completion for reactions with *in vitro* transcribed α-1 and α-2 RNA, whereas the dye-labeled and chemically modified RNAs caused diminished cleavage activity (Figure 4.9B). FnCpf1 appeared the most promising in retaining some activity with dye-labeled and chemically modified RNA; however, α-2 looked like a better targeting sequence compared to α-1. Interestingly, LbCpf1 showed better cleavage with dye-labeled, chemically-modified α-2 in comparison to *in vitro* transcribed α-2. Overall, it is clear that targeting sequence identity (α-1 vs. α-2) has an effect on cleavage activity, and that different Cpf1 species also behave differently, at least in the particular cleavage reaction conditions used here.

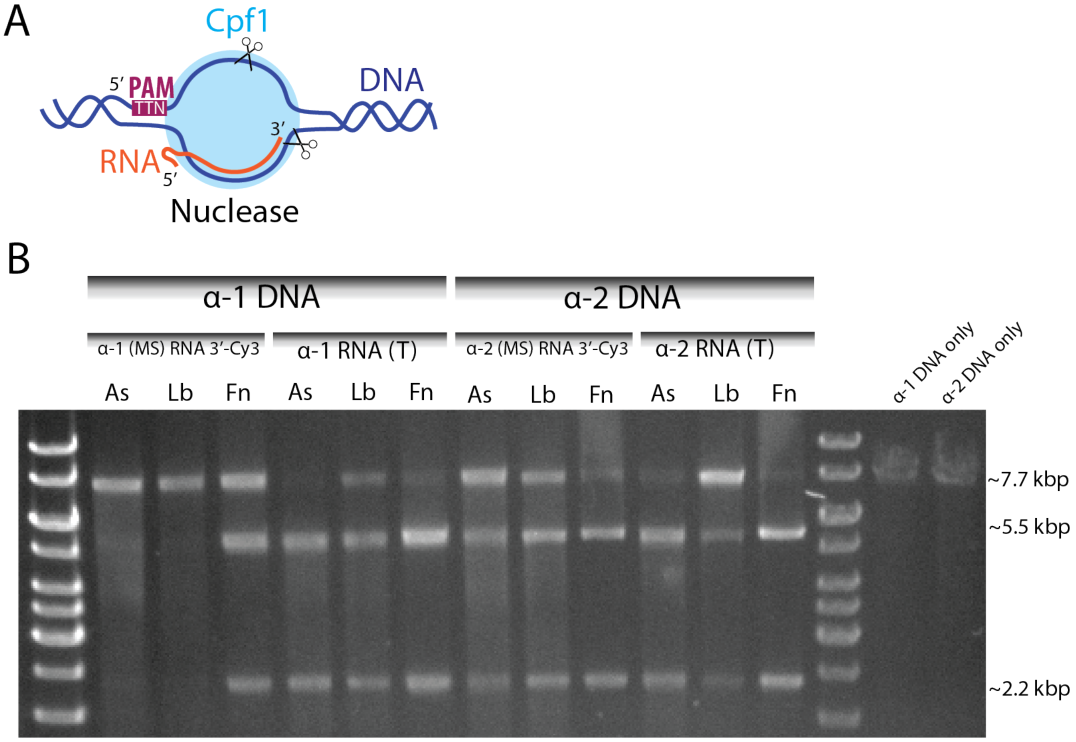


Figure 4.9 *In vitro* cleavage of α-satellite DNA with As, Lb, FnCpf1 and α-1/α-2 RNA containing dye-labels and chemical modifications. Cleavage assay gel showing reactions of DNA cleavage with Cfp1 complexed with α-1 or α-2 containing 3' dye-labels and MS modifications, for comparison, *in vitro* transcribed (T) RNA is used in parallel reactions as a control.

Table 4.5 RNA oligo designs for Cpf1/dCpf1. *(*) phosphorothioate modified backbone; (m) 2'-Omethyl bases. RNAs are purchased from Dharmacon. In gray shading (oligos A1, A2), RNAs were made from in vitro transcription reactions via T7 RNA polymerase.*

Name	**Description**	**Sequence**
α-1	**3′ Cy3 modified TTTN α-1**	5′ mA(*)mA(*)mU(*) UUC UAC UGU UGU AGA UUA GAA UCU GCA AGU GGA UAU U(*)mU(*)mG(*) mG (Cy3) 3′
α-2	**3′ Cy3 modified TTTN α-2**	5′ mA(*)mA(*)mU(*) UUC UAC UGU UGU AGA UUG AUG UGU GCA UUC AAC UCA C(*)mA(*)mG(*) mA (Cy3) 3′
A-1	**TTTN α-1**	5′ GGA UUU CUA CUG UUG UAG AUU AGA AUC UGC AAG UGG AUA UUU GG 3′
A-2	**TTTN α-2**	5′ GGA UUU CUA CUG UUG UAG AUU GAU GUG UGC AUU CAA CUC ACA GA 3′

Discussion

In this work, we show that in *in vitro* cleavage assays, the 2-pc tracrRNA-crRNA can effectively target Cas9 to cleave DNA substrates specifically. Cleavage was just as efficient compared to the 1-pc sgRNA, and in the case of telomere targeting sequence, the 2-pc tracrRNA-crRNA system displayed even higher cleavage activity. Dye-labels and chemical modifications on the RNA oligos showed minimal perturbations to cleavage activity, from which we infer that these modified RNA species also minimally perturbed Cas9's DNA binding activity (ultimately, we are interested in site-specific DNA binding, not cleavage, for visualizing DNA sites using dCas9). Additionally, we tested an alternative CRISPR-Cas system, CRISPR-Cpf1, for possible use in the development of our *in vitro* reconstituted system for DNA visualization. We saw that unlike with Cas9, Cpf1 showed a marked decrease in DNA cleavage activity when the protein is complexed with dye-labeled and chemically modified RNA oligos. The RNA oligos introduced here are currently being used in proof-of-concept experiments in live mammalian cells, where we are optimizing delivery of the *in vitro* reconstituted CRISPR-Cas9 and CRISPR-Cfp1 systems. This project is an on-going effort in our lab.

Materials and Methods

Commercially synthesized RNA oligos

All commercially synthesized RNA oligos were ordered from Dharmacon (dharmacon.gelifesciences.com). All oligos were HPLC purified, desalted and deprotected on arrival. DEPC treated water was used to dissolve the lyophilized powder, and 100 μM - 500 μM stocks were made and stored in small aliquots at -80°C until later use.

In vitro transcription of RNA oligos

When handling RNA, general care was taken to ensure work space and all lab materials and reagents used were RNase free. RNaseZap (Ambion AM9780) was used to wipe down surfaces and containers difficult to decontaminate by other means.

Depending on the length of the desired single stranded RNA product, two approaches were used to preparing the DNA templates for in vitro transcription reactions. For shorter RNA products <100 bases (i.e. crRNA), DNA templates were hybridized commercial DNA oligos which contain a double stranded T7 promoter region. The standard oligo Annealing protocol was used, briefly, both oligos was resuspended in Annealing Buffer (10 mM Tris, pH 7.5-8.0, 50 mM NaCl, 1 mM EDTA), mixed at 1:1 ratio in a 1.5 mL tube, heated to 95^{o}C for 5 minutes, removed from heat and slowly cooled down to RT (around 45 min), the hybridized oligos were store at 4^{o}C until ready to use. For longer RNA products >100 bases (i.e. sgRNA), a plasmid containing the desired sgRNA sequence was used as a template, standard PCR reactions were performed using *Taq* polymerase with Standard *Taq* Buffer (NEB M0273) according to protocol. Primers containing T7 promoter region upstream of the sgRNA sequence was used, resulting in sgRNA DNA templates suitable for T7 RNAP in vitro transcription.

Standard in vitro transcription protocol was as follows: for a 100 uL reaction, with a final concentration of 1X reaction buffer (50 mM Tris-HCL, pH 7.5, 2 mM spermidine, 1 mM dithiothreitol) , 1.5 mM to 15 mM $MgCl_2$, 5 mM each of rNTP (CTP, ATP, GTP, UTP, Thermo Scientific NTP Set 100 mM Solutions R1481), 0.5 μg to 2 μg of DNA template, 2μL of RNAP (homemade from Scott Bailey Lab or NEB T7 RNA Polymerase M0251S), optional inclusion of RNase inhibitor (Thermo Scientific RiboLock RNase Inhibitor EO0381) at 1 U/μL final. H_2O was added to reach the final reaction volume. Typically, 6 identical reactions were run in parallel to increase product yield. Incubation time was for 4 to 6 hours at 37^{o}C. At the end of the incubation time, 5 μL of 500 mM EDTA was added to quench each reaction. The reactions were purified immediately or stored at -20^{o}C until later processing.

A denaturing acrylamide gel was used to purify the single stranded RNA oligo, followed by ethanol precipitation. Typically, an 8 % gel was run while the gel plate was heated to 50°C to prevent formation of secondary structures, RNA solutions were heated to 90°C for 1 minute before loading into wells. RNA bands were visualized via UV shadowing and subsequently cut out, a gel piece containing RNA from a single reaction was put into 450 μL of TE buffer and put on rotary shaker overnight at 4°C for extraction. Next day, TE buffer solution was moved to a new 1.5 mL tube and sodium acetate is added to final concentration of 0.3 M (add 50 μL of 3 M). After mixing, 2X volume of ethanol was added. After vortexing the solution, the tube was frozen in liquid nitrogen for 10 min (or alternatively at -80°C for 2 hours). The frozen solution was spun-down at 15000 rpm at 4°C for 30 min, supernatant was removed, and the visible RNA pellet was rinsed with 95% ethanol. The solution was spun-down again at 15000 rpm at 4°C for 10 min again, supernatant was removed, the pellet was air-dried and later dissolved in an appropriate volume of solvent (DEPC treated H_2O, or THE RNA Storage Solution, Ambion AM7000). Conversely, a commercial kit was also used to purify the RNA (MEGAclear Kit, Ambion AM1908). For commercial MEGAclear kit, protocol from the manufacturer was followed. Purified RNA was stored at -20°C until later use.

Enzymatic 3’ end-labeling of RNA oligos

Single-stranded RNA ligase (T4 RNA ligase 1) was used to 3’ end-label desired RNA oligos previously made using in vitro transcription [242]. For a 28 μL labeling reaction, 10 μL of purified unlabeled RNA was used (~2500 ng/μL), 7 μL of DMSO was added before the solution was heated to 90°C for 5 minutes, then placed on ice. Add 2.8 μL of 10X ligase buffer, 2.8 μL of 10 mM ATP, 1 μL of ssRNA ligase (T4 RNA Ligase 1, NEB M0204S), 0.5 μL of 1mM pCp-dye (Jena Biosciences, NU-1706-CY3). The final pCp-dye concentration was ~18 μM. The labeling reaction was placed at 4°C for 4-5 hours protected from light. After labeling reaction is finished, run denaturing acrylamide gel followed by ethanol precipitation to purify and concentrate labeled RNA. Alternatively, MEGAclear kit could be used to purify labeled RNA. Quantify labeled RNA and estimate labeling efficiency using Nanodrop.

In vitro DNA cleavage experiments with CRISPR-Cas9

First, DNA templates for cleavage reactions were prepared. Since longer templates would allow for easier detection via staining of cleaved DNA products in agarose gel, plasmid backbone was used as starting point for making cleavage templates. Site-directed mutagenesis was performed to introduce correct PAM sequence (NGG) adjacent to DNA sequences targeted by sgRNAs. QuikChange Lightning Site-Directed Mutagenesis Kit (Agilent 210518) was used, online tool provided by Agilent was used to design mutagenesis primer, commercial standard protocol was followed. Mutated regions were confirmed through Sanger sequencing. DNA plasmids were usually linearized using restriction enzymes before being used as cleavage templates for CRISPR-Cas9 cleavage experiments, this is to make the cut easier to interpret since the product is two separate pieces, rather than a linearized fragment. All restriction enzymes were purchased from NEB and used according to manufacturer's protocol. In Brief, 1 μg of plasmid DNA was restricted in 50 μL reaction for 1hr. Linearized DNA was gel-purified by running on a 1% Agarose gel in TAE buffer, the desired band was later extracted using GeneJet Gel Extraction Kit (Thermo K0691).

For *in vitro* cleavage experiment, there was slight variation in the experimental protocol from experiment to experiment, namely, the ratios of RNA, DNA and Cas9 protein used (indicated when relevant), but the general cleavage reaction procedure was as follows. One cleavage reaction was around 20 μL in volume, and consisted of 1X reaction buffer, 300 nM RNA, 49.5 nM Cas9 (NEB, M0386S), 2.25 nM DNA (agarose gel purified linearized plasmid containing targeting sequence). Cas9 and RNA was incubated in a 1.5 mL tube for 10 minutes at RT to allow complex formation before the addition of DNA. After addition of DNA and slight mixing, the reaction was incubated for 1-3 hours at 37°C. Afterwards, 30 minutes of Proteinase K (NEB, P8107S) treatment was performed at 37°C. The entire cleavage reaction solution was run on a 1% agarose gel, and DNA fragments was visualized after staining with 0.01% Ethidium Bromide (v/v) for 30 minutes.

In vitro DNA cleavage experiments with CRISPR-Cpf1

First, DNA templates for cleavage reactions were prepared in the same way as described in the previous section for Cas9, and the proper PAM motif was introduced using site directed mutagenesis. Cleavage reactions were optimized and performed using AsCpf1, LbCpf1 and FnCpf1 expressed and purified by John Mallon (Bailey Lab), and Digvijay Singh (Ha Lab); purified protein preparations were from the period of September to October 2016. Yanbo Wang (rotation student, PMB) assisted in performing the cleavage reactions.

Chapter 5: Conclusions and Future Directions

In the preceding chapters, three separate experimental projects were presented and discussed in detail, namely: the investigation of the spatial organization of transcription in *E. coli* (Chapter 2), the observation of transcription factor mediated DNA looping (Chapter 3), and the *in vitro* characterizations of a novel reconstituted CRISPR-Cas system for imaging genomic sites (Chapter 4). In this concluding chapter, a brief summary of the results from these separate works will be provided, along with a succinct exploration of these project's future directions.

In Chapter 2, we investigated the spatial organization of transcription in *E. coli* by using superresolution imaging techniques to visualize the spatial distribution of various molecular components of transcription inside the cell, including DNA sites, pre-rRNA, RNAP, and an essential transcription factor NusA. In addition, we used two-color imaging to characterize the spatial correlation between the different molecular species of transcription, in order to gain insight into the functional significance of their distributions. We found that RNAP formed a distinct clustered distribution in cells under a number of different growth conditions, and this clustering of RNAP seems to be largely independent of global transcription activity. While pre-rRNA synthesis sites were highly colocalized to RNAP clusters in cells under the rich medium growth condition, these RNAP clusters were retained when we down-regulated rRNA synthesis activity. We saw that the elongation factor NusA also formed a distinct clustered distribution and NusA clusters highly colocalized with RNAP clusters in an rRNA transcription activity independent manner. These results suggest that RNAP within clusters likely are associated with NusA, and since NusA is a well-studied transcription regulator that acts on elongation complexes [133], it is likely that RNAP molecules within clusters are elongation complexes either actively undergoing elongation or stalled. Overall, we also saw a cell length dependence on the positioning of RNAP clusters along the long axis of the cell. As cells elongate, RNAP clusters follow the segregating chromosome and move toward the cell poles. This cell length dependence in cluster positioning is independent of transcription activity. We wanted to see if global chromosome organization played a larger role in the organization and positioning of RNAP clusters. When we used gyrase inhibition to perturbed global chromosomal supercoiling density, we saw that pre-rRNA synthesis levels were not changed, and both RNAP clusters and pre-rRNA clusters were retained but more dispersed. There remained a high level of colocalization between RNAP clusters and pre-rRNA clusters, however, there was a redistribution of both RNAP clusters and pre-rRNA clusters to be located more peripherally in the cell dimensions. This indicates that, overall, supercoiling density plays a more important role in the organization of RNAP clusters compared to high transcription activity of *rrn* operons. So far, our experimental results provide evidence supporting the previously proposed transcription factory model only in relatively fast-growing cells: where RNAP clusters are actively synthesizing pre-rRNA, and *rrn* operons can also have high levels of colocalization with RNAP clusters. But our results also show that RNAP clusters are largely independent of both *rrn* operon activity and the physical presence of *rrn* operons, and RNAP clusters can be redistributed in the cell dimensions when we perturb chromosome organization. We observed that RNAP clusters were associated with both *rrn* operons and other mRNA operons, this spatial colocalization (higher than basal level) seems to be independent of transcription activity. It is important to note that a large portion of operons were not always colocalized with RNAP clusters even for the most highly colocalized *rrnD* operon, which indicates that the colocalization of operons and RNAP clusters is likely very dynamic. RNAP clusters likely are composed of both elongating and stalled RNAP elongation complexes, and are associated with a large number of DNA sites, the underlying set of DNA sites that recruit RNAP and create RNAP clusters under different growth conditions may be different. Others in our lab are actively investigating other characteristics of RNAP clusters to better elucidate their functional significance in *E. coli*. Firstly, we are further investigating the nature of RNAP clusters; for example, by using Fluorescence Recovery After Photobleaching (FRAP) experiments, we are looking at the differential mobility of RNAP molecules that are located inside or outside of RNAP clusters. Secondly, we can directly probe the actual DNA sites associated with RNAP clusters using techniques such as Chromatin Interaction Analysis by Pair-End Tag Sequencing (ChIA-PET); this approach is a more direct way to reveal what DNA sequences contribute to the formation of RNAP clusters.

In Chapter 3, we reported the use of a minimal foot-print DNA marker system to observe looping between two DNA sites separated only 2.3 kb apart in live *E. coli* cells. We were able to estimate the looping frequencies of constructs with different operator characteristics and regulatory regimes. Our measurement of the end-to-end separation of the 2.3kb segment is ~71 nm, which is much smaller than predicted using a standard *in vitro* persistence length of 50 nm for a non-interacting worm like chain model of DNA. This is most likely a result of DNA compaction in the native crowded cellular environment; more specifically DNA compaction can be caused by the binding of many DNA associated proteins (NAPs), and also overall DNA supercoiling. This DNA marker system allows for minimal perturbation to the native chromosomal site (compared to traditional FROS), and the resulting diffraction-limited fluorescent spots enable us to image DNA sites at unprecedented high resolution (~40nm). This minimal DNA marker system has proven to be a reliable way to label any DNA site of interest in our lab; similar ways of labeling DNA sites were also implemented in Chapter 2 and for many other on-going projects in our lab.

Lastly, in Chapter 4, we detailed the initial *in vitro* characterizations of a reconstituted CRISPR-Cas system for imaging genomic DNA sites in live mammalian cells. Using *in vitro* cleavage assays, we assessed the cleavage efficiencies of CRISPR-Cas complexes reconstituted with RNA oligos containing chemical modifications and dye-labels. We saw that for the CRISPR-Cas9 system, chemical modifications and dye labels had minimal effect on the cleavage efficiencies of DNA substrates. But for the CRISPR-Cpf1 system, chemically modified and dye labeled RNA oligos overall caused the lowering of cleavage efficiencies. Others in our lab are actively working on the *in vivo* implementation of this reconstituted system for genomic DNA visualization in live mammalian cells. The successful use of chemically modified and dye-labeled RNA oligos in a reconstituted CRISPR-Cas complex delivery system would provide an unprecedented way to simultaneously visualize a large number of genomic DNA sites in live cells by multiplex bar-coding. Due to this method's relative low cost and ease of implementation, the reconstituted system potentially could provide a reliable way to visualize native DNA sequences in a variety of wildtype cancer or primary cell lines.

References

1. Elowitz, M. B., Surette, M. G., Wolf, P. E., Stock, J. B. & Leibler, S. Protein mobility in the cytoplasm of Escherichia coli. *Journal of Bacteriology* **181,** 197–203 (1999).
2. Losick, R. & Shapiro, L. Changing views on the nature of the bacterial cell: from biochemistry to cytology. *Journal of Bacteriology* **181,** 4143–4145 (1999).
3. Bi, E. F. & Lutkenhaus, J. FtsZ ring structure associated with division in Escherichia coli. *Nature* **354,** 161–164 (1991).
4. Valens, M., Penaud, S., Rossignol, M., Cornet, F. & Boccard, F. Macrodomain organization of the Escherichia coli chromosome. *EMBO J* **23,** 4330–4341 (2004).
5. Espéli, O., Mercier, R. & Boccard, F. DNA dynamics vary according to macrodomain topography in the E. coli chromosome. *Molecular Microbiology* **68,** 1418–1427 (2008).
6. Thiel, A., Valens, M., Vallet-Gely, I., Espéli, O. & Boccard, F. Long-range chromosome organization in E. coli: a site-specific system isolates the Ter macrodomain. *PLoS Genet.* **8,** e1002672 (2012).
7. Cagliero, C., Grand, R. S., Jones, M. B., Jin, D. J. & O'Sullivan, J. M. Genome conformation capture reveals that the Escherichia coli chromosome is organized by replication and transcription. *Nucleic Acids Research* **41,** 6058–6071 (2013).
8. Umbarger, M. A. *et al.* The three-dimensional architecture of a bacterial genome and its alteration by genetic perturbation. *Mol. Cell* **44,** 252–264 (2011).
9. Le, T. B. K., Imakaev, M. V., Mirny, L. A. & Laub, M. T. High-Resolution Mapping of the Spatial Organization of a Bacterial Chromosome. *Science* **342,** 731–734 (2013).
10. Liu, L. F. & Wang, J. C. Supercoiling of the DNA template during transcription. *Proc Natl Acad Sci USA* **84,** 7024–7027 (1987).
11. Hsu, Y.-H., Chung, M.-W. & Li, T.-K. Distribution of gyrase and topoisomerase IV on bacterial nucleoid: implications for nucleoid organization. *Nucleic Acids Research* **34,** 3128–3138 (2006).
12. Tadesse, S. & Graumann, P. L. Differential and dynamic localization of topoisomerases in Bacillus subtilis. *Journal of Bacteriology* **188,** 3002–3011 (2006).
13. Luijsterburg, M. S., White, M. F., van Driel, R. & Dame, R. T. The major architects of chromatin: architectural proteins in bacteria, archaea and eukaryotes. *Critical Reviews in Biochemistry and Molecular Biology* **43,** 393–418 (2008).
14. Dillon, S. C. & Dorman, C. J. bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nat. Rev. Microbiol.* **8,** 185–195

(2010).

15. Wang, X. The two Escherichia coli chromosome arms locate to separate cell halves. *Genes & Development* **20,** 1727–1731 (2006).
16. Nielsen, H. J., Ottesen, J. R., Youngren, B., Austin, S. J. & Hansen, F. G. The Escherichia coli chromosome is organized with the left and right chromosome arms in separate cell halves. *Molecular Microbiology* **62,** 331–338 (2006).
17. Badrinarayanan, A., Lesterlin, C., Reyes-Lamothe, R. & Sherratt, D. The Escherichia coli SMC Complex, MukBEF, Shapes Nucleoid Organization Independently of DNA Replication. *Journal of Bacteriology* **194,** 4669–4676 (2012).
18. Berlatzky, I. A., Rouvinski, A. & Ben-Yehuda, S. Spatial organization of a replicating bacterial chromosome. *Proceedings of the National Academy of Sciences* **105,** 14136–14140 (2008).
19. Webb, C. D. *et al.* Bipolar localization of the replication origin regions of chromosomes in vegetative and sporulating cells of B. subtilis. *Cell* **88,** 667–674 (1997).
20. Violin, J. D., Zhang, J., Tsien, R. Y. & Newton, A. C. A genetically encoded fluorescent reporter reveals oscillatory phosphorylation by protein kinase C. *The Journal of Cell Biology* **161,** 899–909 (2003).
21. Toro, E., Hong, S.-H., Mcadams, H. H. & Shapiro, L. Caulobacter requires a dedicated mechanism to initiate chromosome segregation. *Proceedings of the National Academy of Sciences* **105,** 15435–15440 (2008).
22. Lee, P. S., Lin, D. C.-H., Moriya, S. & Grossman, A. D. Effects of the chromosome partitioning protein Spo0J (ParB) on oriC positioning and replication initiation in Bacillus subtilis. *Journal of Bacteriology* **185,** 1326–1337 (2003).
23. Browning, D. F., Grainger, D. C. & Busby, S. J. Effects of nucleoid-associated proteins on bacterial chromosome structure and gene expression. *Curr. Opin. Microbiol.* **13,** 773–780 (2010).
24. Dorman, C. J. Genome architecture and global gene regulation in bacteria: making progress towards a unified model? *Nat. Rev. Microbiol.* **11,** 349–355 (2013).
25. Harrington, E. W. & Trun, N. J. Unfolding of the bacterial nucleoid both in vivo and in vitro as a result of exposure to camphor. *Journal of Bacteriology* **179,** 2435–2439 (1997).
26. Cabrera, J. E., Cagliero, C., Quan, S., Squires, C. L. & Jin, D. J. Active transcription of rRNA operons condenses the nucleoid in Escherichia coli: examining the effect of transcription on nucleoid structure in the absence of transertion. *Journal of Bacteriology* **191,** 4180–4185 (2009).
27. Jin, D. & Cabrera, J. Coupling the distribution of RNA polymerase to global gene regulation and the dynamic structure of the bacterial nucleoid in Escherichia coli. *Journal of structural biology* **156,** 284–291 (2006).

28. Dekker, J. Capturing Chromosome Conformation. *Science* **295,** 1306–1311 (2002).
29. Straight, A. F., Belmont, A. S., Robinett, C. C. & Murray, A. W. GFP tagging of budding yeast chromosomes reveals that protein-protein interactions can mediate sister chromatid cohesion. *Curr. Biol.* **6,** 1599–1608 (1996).
30. Bertrand, E. *et al.* Localization of ASH1 mRNA particles in living yeast. *Mol. Cell* **2,** 437–445 (1998).
31. Ido Golding, Johan Paulsson, Zawilski, S. M. & Edward C Cox. Real-Time Kinetics of Gene Activity in Individual Bacteria. *Cell* **123,** 1025–1036 (2005).
32. Hocine, S., Raymond, P., Zenklusen, D., Chao, J. A. & Singer, R. H. Single-molecule analysis of gene expression using two-color RNA labeling in live yeast. *Nat Meth* **10,** 119–121 (2013).
33. Baron-Benhamou, J., Gehring, N. H., Kulozik, A. E. & Hentze, M. W. Using the lambdaN peptide to tether proteins to RNAs. *Methods Mol. Biol.* **257,** 135–154 (2004).
34. Lawrence, J. B., Villnave, C. A. & Singer, R. H. Sensitive, high-resolution chromatin and chromosome mapping in situ: presence and orientation of two closely integrated copies of EBV in a lymphoma line. *Cell* **52,** 51–61 (1988).
35. Raj, A. & Tyagi, S. Detection of individual endogenous RNA transcripts in situ using multiple singly labeled probes. *Meth. Enzymol.* **472,** 365–386 (2010).
36. Betzig, E. *et al.* Imaging Intracellular Fluorescent Proteins at Nanometer Resolution. *Science* **313,** 1642–1645 (2006).
37. Rust, M. J., Bates, M. & Zhuang, X. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat Meth* **3,** 793–796 (2006).
38. Schmidt, T. & Schütz, G. J. *Handbook of Single-Molecule Biophysics*. 19–42 (Springer US, 2009). doi:10.1007/978-0-387-76497-9_2
39. Carpentier, A.-S., Torrésani, B., Grossmann, A. & Hénaut, A. Decoding the nucleoid organisation of Bacillus subtilis and Escherichia coli through gene expression data. *BMC Genomics* **6,** 84 (2005).
40. Riva, A., Carpentier, A.-S., Barloy-Hubler, F., Chéron, A. & Hénaut, A. Analyzing stochastic transcription to elucidate the nucleoid's organization. *BMC Genomics* **9,** 125–10 (2008).
41. Junier, I., Hérisson, J. & Képès, F. Genomic Organization of Evolutionarily Correlated Genes in Bacteria: Limits and Strategies. *Journal of Molecular Biology* **419,** 369–386 (2012).
42. Audit, B. & Ouzounis, C. A. From Genes to Genomes: Universal Scale-invariant Properties of Microbial Chromosome Organisation. *Journal of Molecular Biology* **332,** 617–633 (2003).
43. Képès, F. Periodic transcriptional organization of the E.coli genome.

Journal of Molecular Biology **340,** 957–964 (2004).

44. Qian, Z., Dimitriadis, E. K., Edgar, R., Eswaramoorthy, P. & Adhya, S. Galactose repressor mediated intersegmental chromosomal connections in Escherichia coli. *Proceedings of the National Academy of Sciences* **109,** 11336–11341 (2012).
45. Iborra, F. J., Pombo, A., Jackson, D. A. & Cook, P. R. Active RNA polymerases are localized within discrete transcription ‘factories’ in human nuclei (vol 109, pg 1427, 1996). *J. Cell. Sci.* **111,** 2280–2280 (1998).
46. Sutherland, H. & Bickmore, W. A. Transcription factories: gene expression in unions? *Nat. Rev. Genet.* **10,** 457–466 (2009).
47. Osborne, C. S. *et al.* Active genes dynamically colocalize to shared sites of ongoing transcription. *Nat Genet* **36,** 1065–1071 (2004).
48. Osborne, C. S. *et al.* Myc Dynamically and Preferentially Relocates to a Transcription Factory Occupied by Igh. *PLoS Biol* **5,** e192 (2007).
49. Ragoczy, T. The locus control region is required for association of the murine beta-globin locus with engaged transcription factories during erythroid maturation. *Genes & Development* **20,** 1447–1457 (2006).
50. Cisse, I. I. *et al.* Real-Time Dynamics of RNA Polymerase II Clustering in Live Human Cells. *Science* **341,** 664–667 (2013).
51. Zhao, Z. W. *et al.* Spatial organization of RNA polymerase II inside a mammalian cell nucleus revealed by reflected light-sheet superresolution microscopy. *Proceedings of the National Academy of Sciences* (2013). doi:10.1073/pnas.1318496111
52. Cabrera, J. E. & Jin, D. J. The distribution of RNA polymerase in Escherichia coli is dynamic and sensitive to environmental cues. *Molecular Microbiology* **50,** 1493–1505 (2003).
53. Cabrera, J. E. & Jin, D. J. Active transcription of rRNA operons is a driving force for the distribution of RNA polymerase in bacteria: effect of extrachromosomal copies of rrnB on the in vivo localization of RNA polymerase. *Journal of Bacteriology* **188,** 4007–4014 (2006).
54. Lewis, P. J., Thaker, S. D. & Errington, J. Compartmentalization of transcription and translation in Bacillus subtilis. *EMBO J* **19,** 710–718 (2000).
55. Bremer, H., Berry, L. & Dennis, P. P. Regulation of ribonucleic acid synthesis in Escherichia coli B-r: an analysis of a shift-up. II. Fraction of RNA polymerase engaged in the synthesis of stable RNA at different steady-state growth rates. *Journal of Molecular Biology* **75,** 161–179 (1973).
56. Bremer, H. & Dennis, P. P. Modulation of chemical composition and other parameters of the cell by growth rate. *Escherichia coli and Salmonella: cellular and molecular biology* **2,** 1553–1569 (1996).
57. Klumpp, S. & Hwa, T. Growth-rate-dependent partitioning of RNA polymerases in bacteria. *Proceedings of the National Academy of*

Sciences **105,** 20245–20250 (2008).

58. Berlyn, M. K. Linkage map of Escherichia coli K-12, edition 10: the traditional map. *Microbiology and Molecular Biology Reviews* **62,** 814–984 (1998).
59. Piggot, P. J. & Hoch, J. A. Revised genetic linkage map of Bacillus subtilis. *Microbiol. Rev.* **49,** 158–179 (1985).
60. Wang, W., Li, G.-W., Chen, C., Xie, X. S. & Zhuang, X. Chromosome organization by a nucleoid-associated protein in live bacteria. *Science* **333,** 1445–1449 (2011).
61. de Wit, E. & de Laat, W. A decade of 3C technologies: insights into nuclear organization. *Genes & Development* **26,** 11–24 (2012).
62. Umbarger, M. A. Chromosome conformation capture assays in bacteria. *Methods* **58,** 212–220 (2012).
63. Ryter, A. & Chang, A. Localization of transcribing genes in the bacterial cell by means of high resolution autoradiography. *Journal of Molecular Biology* **98,** 797–810 (1975).
64. Krawiec, S. & Riley, M. Organization of the bacterial chromosome. *Microbiol. Rev.* **54,** 502–539 (1990).
65. Papantonis, A. *et al.* Active RNA Polymerases: Mobile or Immobile Molecular Machines? *PLoS Biol* **8,** e1000419 (2010).
66. Sánchez Romero, M. A., Lee, D. J., Sánchez Morán, E. & Busby, S. J. W. Location and dynamics of an active promoter in Escherichia coliK-12. *Biochem. J.* **441,** 481–485 (2011).
67. Libby, E. A., Roggiani, M. & Goulian, M. Membrane protein expression triggers chromosomal locus repositioning in bacteria. *Proceedings of the National Academy of Sciences* **109,** 7445–7450 (2012).
68. Mohl, D. A. & Gober, J. W. Cell cycle-dependent polar localization of chromosome partitioning proteins in Caulobacter crescentus. *Cell* **88,** 675–684 (1997).
69. Bates, D. & Kleckner, N. Chromosome and replisome dynamics in E. coli: loss of sister cohesion triggers global chromosome movement and mediates chromosome segregation. *Cell* **121,** 899–911 (2005).
70. Lau, I. F. *et al.* Spatial and temporal organization of replicating Escherichia coli chromosomes. *Molecular Microbiology* **49,** 731–743 (2004).
71. Wang, X., Possoz, C. & Sherratt, D. J. Dancing around the divisome: asymmetric chromosome segregation in Escherichia coli. *Genes & Development* **19,** 2367–2377 (2005).
72. Viollier, P. H. *et al.* Rapid and sequential movement of individual chromosomal loci to specific subcellular locations during bacterial DNA replication. *Proc Natl Acad Sci USA* **101,** 9257–9262 (2004).
73. Wiggins, P. A., Cheveralls, K. C., Martin, J. S., Lintner, R. & Kondev, J. Strong intranucleoid interactions organize the Escherichia coli chromosome into a nucleoid filament. *Proceedings of the National*

Academy of Sciences **107,** 4991–4995 (2010).

74. Hong, S.-H. *et al.* Caulobacter chromosome in vivo configuration matches model predictions for a supercoiled polymer in a cell-like confinement. *Proc Natl Acad Sci USA* **110,** 1674–1679 (2013).

75. Mahy, N. L. Spatial organization of active and inactive genes and noncoding DNA within chromosome territories. *The Journal of Cell Biology* **157,** 579–589 (2002).

76. Hensel, Z., Weng, X., Lagda, A. C. & Xiao, J. Transcription-Factor-Mediated DNA Looping Probed by High-Resolution, Single-Molecule Imaging in Live E. coli Cells. *PLoS Biol* **11,** e1001591 (2013).

77. Thompson, R. E., Larson, D. R. & Webb, W. W. Precise nanometer localization analysis for individual fluorescent probes. *Biophysj* **82,** 2775–2783 (2002).

78. Elf, J., Li, G.-W. & Xie, X. S. Probing transcription factor dynamics at the single-molecule level in a living cell. *Science* **316,** 1191–1194 (2007).

79. Kuhlman, T. E. & Cox, E. C. Gene location and DNA density determine transcription factor distributions in Escherichia coli. *Mol. Syst. Biol.* **8,** 1–13 (2012).

80. Bintu, L. *et al.* Transcriptional regulation by the numbers: applications. *Current Opinion in Genetics & Development* **15,** 125–135 (2005).

81. Janga, S. C., Salgado, H. & Martinez-Antonio, A. Transcriptional regulation shapes the organization of genes on bacterial chromosomes. *Nucleic Acids Research* **37,** 3680–3688 (2009).

82. Smits, W. K. & Grossman, A. D. The transcriptional regulator Rok binds A+T-rich DNA and is involved in repression of a mobile genetic element in Bacillus subtilis. *PLoS Genet.* **6,** e1001207 (2010).

83. Lee, S. F., Thompson, M. A., Schwartz, M. A., Shapiro, L. & Moerner, W. E. Super-Resolution Imaging of the Nucleoid-Associated Protein HU in Caulobacter crescentus. *Biophysj* **100,** L31–L33 (2011).

84. Geanacopoulos, M. & Adhya, S. Genetic analysis of GalR tetramerization in DNA looping during repressosome assembly. *J. Biol. Chem.* **277,** 33148–33152 (2002).

85. Ueguchi, C., Seto, C., Suzuki, T. & Mizuno, T. Clarification of the dimerization domain and its functional significance for the Escherichia coli nucleoid protein H-NS. *Journal of Molecular Biology* **274,** 145–151 (1997).

86. Lim, C. J., Lee, S. Y., Kenney, L. J. & Yan, J. Nucleoprotein filament formation is the structural basis for bacterial protein H-NS gene silencing. *Sci Rep* **2,** 509 (2012).

87. Azam, T. A. & Ishihama, A. Twelve species of the nucleoid-associated protein from Escherichia coli sequence recognition specificity and DNA binding affinity. *Journal of Biological Chemistry* **274,** 33105–33113 (1999).

88. Sobetzko, P., Travers, A. & Muskhelishvili, G. Gene order and

chromosome dynamics coordinate spatiotemporal gene expression during the bacterial growth cycle. *Proceedings of the National Academy of Sciences* **109,** E42–50 (2012).
89. Lewis, D. E., Geanacopoulos, M. & Adhya, S. Role of HU and DNA supercoiling in transcription repression: specialized nucleoprotein repression complex at gal promoters in Escherichia coli. *Molecular Microbiology* **31,** 451–461 (1999).
90. Cui, L., Murchland, I., Shearwin, K. E. & Dodd, I. B. Enhancer-like long-range transcriptional activation by λ CI-mediated DNA looping. *Proceedings of the National Academy of Sciences* **110,** 2922–2927 (2013).
91. Fritsche, M., Li, S., Heermann, D. W. & Wiggins, P. A. A model for Escherichia coli chromosome packaging supports transcription factor-induced DNA domain formation. *Nucleic Acids Research* **40,** 972–980 (2012).
92. Ebright, R. H. RNA Polymerase: Structural Similarities Between Bacterial RNA Polymerase and Eukaryotic RNA Polymerase II. *Journal of Molecular Biology* **304,** 687–698 (2000).
93. Gruber, T. M. & Gross, C. A. Assay of Escherichia coli RNA polymerase: sigma-core interactions. *Meth. Enzymol.* **370,** 206–212 (2003).
94. Shepherd, N., Dennis, P. & Bremer, H. Cytoplasmic RNA Polymerase in Escherichia coli. *Journal of Bacteriology* **183,** 2527–2534 (2001).
95. Endesfelder, U. *et al.* Multiscale Spatial Organization of RNA Polymerase in Escherichia coli. *Biophysj* **105,** 172–181 (2013).
96. Bakshi, S., Siryaporn, A., Goulian, M. & Weisshaar, J. C. Superresolution imaging of ribosomes and RNA polymerase in live Escherichia coli cells. *Molecular Microbiology* no–no (2012). doi:10.1111/j.1365-2958.2012.08081.x
97. Grainger, D. C., Hurd, D., Harrison, M., Holdstock, J. & Busby, S. J. W. Studies of the distribution of Escherichia coli cAMP-receptor protein and RNA polymerase along the E. coli chromosome. *Proc Natl Acad Sci USA* **102,** 17693–17698 (2005).
98. Herring, C. D. *et al.* Immobilization of Escherichia coli RNA polymerase and location of binding sites by use of chromatin immunoprecipitation and microarrays. *Journal of Bacteriology* **187,** 6166–6174 (2005).
99. Bremer, H., Dennis, P. & Ehrenberg, M. Free RNA polymerase and modeling global transcription in Escherichia coli. *Biochimie* **85,** 597–609 (2003).
100. Shepherd, N. S., Churchward, G. & Bremer, H. Synthesis and activity of ribonucleic acid polymerase in Escherichia coli. *Journal of Bacteriology* **141,** 1098–1108 (1980).
101. Bratton, B. P., Mooney, R. A. & Weisshaar, J. C. Spatial distribution and diffusive motion of RNA polymerase in live Escherichia coli. *Journal of Bacteriology* **193,** 5138–5146 (2011).

102. Bakshi, S., Dalrymple, R. M., Li, W., Choi, H. & Weisshaar, J. C. Partitioning of RNA polymerase activity in live Escherichia coli from analysis of single-molecule diffusive trajectories. *Biophys. J.* **105,** 2676–2686 (2013).

103. Jao, C. Y. & Salic, A. Exploring RNA transcription and turnover in vivo by using click chemistry. *Proc Natl Acad Sci USA* **105,** 15779–15784 (2008).

104. Burmann, B. M. & Rösch, P. The role of E. coli Nus-Factors in transcription regulation and transcription:translation coupling: From structure to mechanism. *Transcription* **2,** 130–134 (2011).

105. Robinow, C. & Kellenberger, E. The bacterial nucleoid revisited. *Microbiol. Rev.* **58,** 211–232 (1994).

106. Briegel, A. *et al.* Multiple large filament bundles observed in Caulobacter crescentus by electron cryotomography. *Molecular Microbiology* **62,** 5–14 (2006).

107. Montero Llopis, P. *et al.* Spatial organization of the flow of genetic information in bacteria. *Nature* **466,** 77–81 (2010).

108. Miller, O. L., Hamkalo, B. A. & Thomas, C. A. Visualization of bacterial genes in action. *Science* **169,** 392–395 (1970).

109. French, S. L. & Miller, O. L. Transcription mapping of the Escherichia coli chromosome by electron microscopy. *Journal of Bacteriology* **171,** 4207–4216 (1989).

110. Cagliero, C. & Jin, D. J. Dissociation and re-association of RNA polymerase with DNA during osmotic stress response in Escherichia coli. *Nucleic Acids Research* (2012). doi:10.1093/nar/gks988

111. English, B. P. *et al.* Single-molecule investigations of the stringent response machinery in living bacterial cells. *Proceedings of the National Academy of Sciences* **108,** E365–73 (2011).

112. Golding, I. & Cox, E. C. RNA dynamics in live Escherichia coli cells. *Proc Natl Acad Sci USA* **101,** 11310–11315 (2004).

113. Valencia-Burton, M., McCullough, R. M., Cantor, C. R. & Broude, N. E. RNA visualization in live bacterial cells using fluorescent protein complementation. *Nat Meth* **4,** 421–427 (2007).

114. Driessen, A. J. M. & Nouwen, N. Protein translocation across the bacterial cytoplasmic membrane. *Annu. Rev. Biochem.* **77,** 643–667 (2008).

115. Kawamoto, H., Morita, T., Shimizu, A., Inada, T. & Aiba, H. Implication of membrane localization of target mRNA in the action of a small RNA: mechanism of post-transcriptional regulation of glucose transporter in Escherichia coli. *Genes & Development* **19,** 328–338 (2005).

116. Nevo-Dinur, K., Nussbaum-Shochat, A., Ben-Yehuda, S. & Amster-Choder, O. Translation-Independent Localization of mRNA in E. coli. *Science* **331,** 1081–1084 (2011).

117. Mackie, G. A. RNase E: at the interface of bacterial RNA processing and decay. *Nature Publishing Group* **11,** 45–57 (2013).

118. Liou, G. G., Jane, W. N., Cohen, S. N., Lin, N. S. & Lin-Chao, S. RNA degradosomes exist in vivo in Escherichia coli as multicomponent complexes associated with the cytoplasmic membrane via the N-terminal region of ribonuclease E. *Proc Natl Acad Sci USA* **98,** 63–68 (2001).
119. Taghbalout, A. & Rothfield, L. RNaseE and the other constituents of the RNA degradosome are components of the bacterial cytoskeleton. *Proc Natl Acad Sci USA* **104,** 1667–1672 (2007).
120. Russell, J. H. & Keiler, K. C. Subcellular localization of a bacterial regulatory RNA. *Proceedings of the National Academy of Sciences* **106,** 16405–16409 (2009).
121. Keiler, K. C. Biology of trans-translation. *Annu. Rev. Microbiol.* **62,** 133–151 (2008).
122. Stracy, M. *et al.* Live-cell superresolution microscopy reveals the organization of RNA polymerase in the bacterial nucleoid. *Proceedings of the National Academy of Sciences* **112,** E4390–9 (2015).
123. Cook, P. R. A model for all genomes: the role of transcription factories. *Journal of Molecular Biology* **395,** 1–10 (2010).
124. Marenduzzo, D., Faro-Trindade, I. & Cook, P. What are the molecular ties that maintain genomic loops? *TRENDS in Genetics* **23,** 126–133 (2007).
125. Melnik, S. *et al.* The proteomes of transcription factories containing RNA polymerases I, II or III. *Nat Meth* **8,** 963–968 (2011).
126. Eskiw, C. H., Rapp, A., Carter, D. R. F. & Cook, P. R. RNA polymerase II activity is located on the surface of protein-rich transcription factories. *J. Cell. Sci.* **121,** 1999–2007 (2008).
127. Gaal, T. *et al.* Colocalization of distant chromosomal loci in space in E. coli: a bacterial nucleolus. *Genes & Development* **30,** 2272–2285 (2016).
128. Stracy, M. & Kapanidis, A. N. Single-molecule and super-resolution imaging of transcription in living bacteria. *Methods* 1–12 (2017). doi:10.1016/j.ymeth.2017.04.001
129. Wang, S., Moffitt, J. R., Dempsey, G. T., Xie, X. S. & Zhuang, X. Characterization and development of photoactivatable fluorescent proteins for single-molecule-based superresolution imaging. *Proceedings of the National Academy of Sciences* **111,** 8452–8457 (2014).
130. Severinova, E. E. A. Inhibition of Escherichia coli RNA Polymerase by Bacteriophage T4 AsiA. 1–10 (1998).
131. Campbell, E. A. *et al.* Structural mechanism for rifampicin inhibition of bacterial rna polymerase. *Cell* **104,** 901–912 (2001).
132. Malagon, F. RNase III is required for localization to the nucleoid of the 5' pre-rRNA leader and for optimal induction of rRNA synthesis in E. coli. *RNA* **19,** 1200–1207 (2013).
133. Mooney, R. A. *et al.* Regulator trafficking on bacterial transcription units in vivo. *Mol. Cell* **33,** 97–108 (2009).
134. Gourse, R. L., Gaal, T., Bartlett, M. S., Appleman, J. A. & Ross, W. rRNA transcription and growth rate-dependent regulation of ribosome synthesis

in Escherichia coli. *Annu. Rev. Microbiol.* **50,** 645–677 (1996).

135. Torres, M., Condon, C., Balada, J. M., Squires, C. & Squires, C. L. Ribosomal protein S4 is a transcription factor with properties remarkably similar to NusA, a protein involved in both non-ribosomal and ribosomal RNA antitermination. *EMBO J* **20,** 3811–3820 (2001).
136. Vogel, U. & Jensen, K. F. NusA is required for ribosomal antitermination and for modulation of the transcription elongation rate of both antiterminated RNA and mRNA. *J. Biol. Chem.* **272,** 12265–12271 (1997).
137. Yang, X. *et al.* The structure of bacterial RNA polymerase in complex with the essential transcription elongation factor NusA. *EMBO Rep.* **10,** 997–1002 (2009).
138. Schweimer, K. *et al.* NusA interaction with the α subunit of E. coli RNA polymerase is via the UP element site and releases autoinhibition. *Structure* **19,** 945–954 (2011).
139. Artsimovitch, I. & Landick, R. Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. *Proc Natl Acad Sci USA* **97,** 7090–7095 (2000).
140. Durfee, T., Hansen, A.-M., Zhi, H., Blattner, F. R. & Jin, D. J. Transcription profiling of the stringent response in Escherichia coli. *Journal of Bacteriology* **190,** 1084–1096 (2008).
141. Hauryliuk, V., Atkinson, G. C., Murakami, K. S., Tenson, T. & Gerdes, K. Recent functional insights into the role of (p)ppGpp in bacterial physiology. *Nature Publishing Group* **13,** 298–309 (2015).
142. Bremer, H. & Ehrenberg, M. Guanosine tetraphosphate as a global regulator of bacterial RNA synthesis: a model involving RNA polymerase pausing and queuing. *Biochim. Biophys. Acta* **1262,** 15–36 (1995).
143. Wahle, E. & Mueller, K. Involvement of DNA gyrase in rRNA synthesis in vivo. *Mol. Gen. Genet.* **179,** 661–667 (1980).
144. Nöllmann, M., Crisona, N. J. & Arimondo, P. B. Thirty years of Escherichia coli DNA gyrase: From in vivo function to single-molecule mechanism. *Biochimie* **89,** 490–499 (2007).
145. Chong, S., Chen, C., Ge, H. & Xie, X. S. Mechanism of transcriptional bursting in bacteria. *Cell* **158,** 314–326 (2014).
146. Rovinskiy, N., Agbleke, A. A., Chesnokova, O., Pang, Z. & Higgins, N. P. Rates of Gyrase Supercoiling and Transcription Elongation Control Supercoil Density in a Bacterial Chromosome. *PLoS Genet.* **8,** e1002845–15 (2012).
147. Peter, B. J. *et al.* Genomic transcriptional response to loss of chromosomal supercoiling in Escherichia coli. *Genome Biol.* **5,** R87 (2004).
148. Higgins, C. F., Dorman, C. J., Stirling, D. A. & Waddell, L. A physiological role for DNA supercoiling in the osmotic regulation of gene expression in S. typhimurium and E. coli. *Cell* **52,** 569–584 (1988).

149. Alt, S., Mitchenall, L. A., Maxwell, A. & Heide, L. Inhibition of DNA gyrase and DNA topoisomerase IV of Staphylococcus aureus and Escherichia coli by aminocoumarin antibiotics. *Journal of Antimicrobial Chemotherapy* **66,** 2061–2069 (2011).
150. Collin, F., Karkare, S. & Maxwell, A. Exploiting bacterial DNA gyrase as a drug target: current state and perspectives. *Appl. Microbiol. Biotechnol.* **92,** 479–497 (2011).
151. Thanbichler, M., Wang, S. C. & Shapiro, L. The bacterial nucleoid: a highly organized and dynamic structure. *Journal of cellular biochemistry* **96,** 506–521 (2005).
152. Higgins, N. P. Species-specific supercoil dynamics of the bacterial nucleoid. *Biophysical Reviews* 1–9 (2016). doi:10.1007/s12551-016-0207-9
153. Sage, D. *et al.* Quantitative evaluation of software packages for single-molecule localization microscopy. *Nat Meth* **12,** 717–724 (2015).
154. Skinner, S. O., lveda, L. A. S. U., Xu, H. & Golding, I. nprot.2013.066. *Nature Protocols* **8,** 1100–1113 (2013).
155. Dempsey, G. T., Vaughan, J. C., Chen, K. H., Bates, M. & Zhuang, X. Evaluation of fluorophores for optimal performance in localization-based super-resolution imaging. *Nat Meth* **8,** 1027–1036 (2011).
156. Schleif, R. DNA looping. *Annu. Rev. Biochem.* **61,** 199–223 (1992).
157. So, L.-H. *et al.* General properties of transcriptional time series in Escherichia coli. *Nature Publishing Group* **43,** 554–560 (2011).
158. Hensel, Z. *et al.* Stochastic expression dynamics of a transcription factor revealed by single-molecule noise analysis. *Nat. Struct. Mol. Biol.* **19,** 797–802 (2012).
159. Salman, H. *et al.* Universal protein fluctuations in populations of microorganisms. *Phys. Rev. Lett.* **108,** 238105 (2012).
160. Choi, P. J., Cai, L., Frieda, K. & Xie, X. S. A stochastic single-molecule event triggers phenotype switching of a bacterial cell. *Science* **322,** 442–446 (2008).
161. Vilar, J. M. G. & Leibler, S. DNA looping and physical constraints on transcription regulation. *Journal of Molecular Biology* **331,** 981–989 (2003).
162. Dunn, T. M., Hahn, S., Ogden, S. & Schleif, R. F. An operator at -280 base pairs that is required for repression of araBAD operon promoter: addition of DNA helical turns between the operator and promoter cyclically hinders repression. *Proc Natl Acad Sci USA* **81,** 5017–5020 (1984).
163. Müller-Hill, B. The function of auxiliary operators. *Molecular Microbiology* **29,** 13–18 (1998).
164. Dandanell, G., Valentin-Hansen, P., Larsen, J. E. & Hammer, K. Long-range cooperativity between gene regulatory sequences in a prokaryote. *Nature* **325,** 823–826 (1987).

165. Wyman, C., Rombel, I., North, A. K., Bustamante, C. & Kustu, S. Unusual oligomerization required for activity of NtrC, a bacterial enhancer-binding protein. *Science* **275,** 1658–1661 (1997).
166. Révet, B., Wilcken-Bergmann, von, B., Bessert, H., Barker, A. & Müller-Hill, B. Four dimers of lambda repressor bound to two suitably spaced pairs of lambda operators form octamers and DNA loops over large distances. *Curr. Biol.* **9,** 151–154 (1999).
167. Dodd, I. B., Perkins, A. J., Tsemitsidis, D. & Egan, J. B. Octamerization of lambda CI repressor is needed for effective repression of P(RM) and efficient switching from lysogeny. *Genes & Development* **15,** 3013–3022 (2001).
168. Simonis, M., Kooren, J. & de Laat, W. An evaluation of 3C-based methods to capture DNA interactions. *Nat Meth* **4,** 895–901 (2007).
169. Dekker, J. The three ‘C’ s of chromosome conformation capture: controls, controls, controls. *Nat Meth* **3,** 17–21 (2006).
170. Houston, P. L. & Broach, J. R. The dynamics of homologous pairing during mating type interconversion in budding yeast. *PLoS Genet.* **2,** e98 (2006).
171. Possoz, C., Filipe, S. R., Grainge, I. & Sherratt, D. J. Tracking of controlled Escherichia coli replication fork stalling and restart at repressor-bound DNA in vivo. *EMBO J* **25,** 2596–2604 (2006).
172. Churchman, L. S., Okten, Z., Rock, R. S., Dawson, J. F. & Spudich, J. A. Single molecule high-resolution colocalization of Cy3 and Cy5 attached to macromolecules measures intramolecular distances through time. *Proc Natl Acad Sci USA* **102,** 1419–1423 (2005).
173. Rozanov, D. V., D'Ari, R. & Sineoky, S. P. RecA-independent pathways of lambdoid prophage induction in Escherichia coli. *Journal of Bacteriology* **180,** 6306–6315 (1998).
174. Little, J. W., Shepley, D. P. & Wert, D. W. Robustness of a gene regulatory circuit. *EMBO J* **18,** 4299–4307 (1999).
175. Aurell, E., Brown, S., Johanson, J. & Sneppen, K. Stability puzzles in phage λ. *Phys. Rev. E* **65,** 2185–9 (2002).
176. Amir, A., Kobiler, O., Rokney, A., Oppenheim, A. B. & Stavans, J. Noise in timing and precision of gene activities in a genetic cascade. *Mol. Syst. Biol.* **3,** 71 (2007).
177. Atsumi, S. & Little, J. W. Regulatory circuit design and evolution using phage lambda. *Genes & Development* **18,** 2086–2094 (2004).
178. Michalowski, C. B., Short, M. D. & Little, J. W. Sequence tolerance of the phage lambda PRM promoter: implications for evolution of gene regulatory circuitry. *Journal of Bacteriology* **186,** 7988–7999 (2004).
179. Becskei, A. & Serrano, L. Engineering stability in gene networks by autoregulation. *Nature* **405,** 590–593 (2000).
180. Michalowski, C. B. & Little, J. W. Positive autoregulation of cI is a dispensable feature of the phage lambda gene regulatory circuitry.

Journal of Bacteriology **187,** 6430–6442 (2005).
181. Johnson, A. D., Meyer, B. J. & Ptashne, M. Interactions between DNA-bound repressors govern regulation by the lambda phage repressor. *Proc Natl Acad Sci USA* **76,** 5061–5065 (1979).
182. Johnson, A. D. *et al.* lambda Repressor and cro--components of an efficient molecular switch. *Nature* **294,** 217–223 (1981).
183. Shea, M. A. & Ackers, G. K. The OR control system of bacteriophage lambda. A physical-chemical model for gene regulation. *Journal of Molecular Biology* **181,** 211–230 (1985).
184. Babić, A. C. & Little, J. W. Cooperative DNA binding by CI repressor is dispensable in a phage lambda variant. *Proc Natl Acad Sci USA* **104,** 17741–17746 (2007).
185. Dodd, I. B. *et al.* Cooperativity in long-range gene regulation by the lambda CI repressor. *Genes & Development* **18,** 344–354 (2004).
186. Reichardt, L. & Kaiser, A. D. Control of lambda repressor synthesis. *Proc Natl Acad Sci USA* **68,** 2185–2189 (1971).
187. Meyer, B. J. & Ptashne, M. Gene regulation at the right operator (OR) of bacteriophage lambda. III. lambda repressor directly activates gene transcription. *Journal of Molecular Biology* **139,** 195–205 (1980).
188. Hawley, D. K. & McClure, W. R. Mechanism of activation of transcription initiation from the lambda PRM promoter. *Journal of Molecular Biology* **157,** 493–525 (1982).
189. Wang, H., Finzi, L., Lewis, D. E. A. & Dunlap, D. AFM studies of lambda repressor oligomers securing DNA loops. *Curr Pharm Biotechnol* **10,** 494–501 (2009).
190. Zurla, C. *et al.* Novel tethered particle motion analysis of CI protein-mediated DNA looping in the regulation of bacteriophage lambda. *J Phys Condens Matter* **18,** S225–S234 (2006).
191. Zurla, C. *et al.* Direct demonstration and quantification of long-range DNA looping by the lambda bacteriophage repressor. *Nucleic Acids Research* **37,** 2789–2795 (2009).
192. Finzi, L. & Dunlap, D. D. Single-molecule approaches to probe the structure, kinetics, and thermodynamics of nucleoprotein complexes that regulate transcription. *Journal of Biological Chemistry* **285,** 18973–18978 (2010).
193. Manzo, C., Zurla, C., Dunlap, D. D. & Finzi, L. The effect of nonspecific binding of lambda repressor on DNA looping dynamics. *Biophys. J.* **103,** 1753–1761 (2012).
194. Sadler, J. R., Sasmor, H. & Betz, J. L. A perfectly symmetric lac operator binds the lac repressor very tightly. *Proc Natl Acad Sci USA* **80,** 6785–6789 (1983).
195. Stirling Churchman, L., Flyvbjerg, H. & Spudich, J. A. A Non-Gaussian Distribution Quantifies Distances Measured with Fluorescence Localization Techniques. *Biophys. J.* **90,** 668–671 (2006).

196. Lu, Y., Weers, B. & Stellwagen, N. C. DNA persistence length revisited. *Biopolymers* **61,** 261–275 (2001).
197. Thanbichler, M. & Shapiro, L. Chromosome organization and segregation in bacteria. *Journal of structural biology* **156,** 292–303 (2006).
198. O'Sullivan, J. M. *et al.* Gene loops juxtapose promoters and terminators in yeast. *Nat Genet* **36,** 1014–1018 (2004).
199. Tan-Wong, S. M., French, J. D., Proudfoot, N. J. & Brown, M. A. Dynamic interactions between the promoter and terminator regions of the mammalian BRCA1 gene. *Proceedings of the National Academy of Sciences* **105,** 5160–5165 (2008).
200. Sarai, A. & Takeda, Y. Lambda repressor recognizes the approximately 2-fold symmetric half-operator sequences asymmetrically. *Proc Natl Acad Sci USA* **86,** 6513–6517 (1989).
201. Beckett, D., Burz, D. S., Ackers, G. K. & Sauer, R. T. Isolation of .lambda. repressor mutants with defects in cooperative operator binding. *Biochemistry* **32,** 9073–9079 (2002).
202. Burz, D. S. & Ackers, G. K. Cooperativity mutants of bacteriophage lambda cI repressor: temperature dependence of self-assembly. *Biochemistry* **35,** 3341–3350 (1996).
203. Stayrook, S., Jaru-Ampornpan, P., Ni, J., Hochschild, A. & Lewis, M. Crystal structure of the λ repressor and a model for pairwise cooperative operator binding. *Nature* **452,** 1022–1025 (2008).
204. Shimada, J. & Yamakawa, H. Ring-closure probabilities for twisted wormlike chains. Application to DNA. *Macromolecules* **17,** 689–698 (1984).
205. Becker, N. B., Rosa, A. & Everaers, R. The radial distribution function of worm-like chains. *Eur Phys J E Soft Matter* **32,** 53–69 (2010).
206. Vafabakhsh, R. & Ha, T. Extreme bendability of DNA less than 100 base pairs long revealed by single-molecule cyclization. *Science* **337,** 1097–1101 (2012).
207. van Noort, J., Verbrugge, S., Goosen, N., Dekker, C. & Dame, R. T. Dual architectural roles of HU: formation of flexible hinges and rigid filaments. *Proc Natl Acad Sci USA* **101,** 6969–6974 (2004).
208. Ringrose, L., Chabanis, S., Angrand, P. O., Woodroofe, C. & Stewart, A. F. Quantitative comparison of DNA looping in vitro and in vivo: chromatin increases effective DNA flexibility at short distances. *EMBO J* **18,** 6630–6641 (1999).
209. Swinger, K. K. & Rice, P. A. IHF and HU: flexible architects of bent DNA. *Curr. Opin. Struct. Biol.* **14,** 28–35 (2004).
210. Vologodskii, A. V., Levene, S. D., Klenin, K. V., Frank-Kamenetskii, M. & Cozzarelli, N. R. Conformational and thermodynamic properties of supercoiled DNA. *Journal of Molecular Biology* **227,** 1224–1243 (1992).
211. Lanctôt, C., Cheutin, T., Cremer, M., Cavalli, G. & Cremer, T. Dynamic genome architecture in the nuclear space: regulation of gene expression

in three dimensions. *Nat. Rev. Genet.* **8,** 104–115 (2007).
212. Janga, S. C., Collado-Vides, J. & Babu, M. M. Transcriptional regulation constrains the organization of genes on eukaryotic chromosomes. *Proc Natl Acad Sci USA* **105,** 15761–15766 (2008).
213. Saad, H. *et al.* DNA dynamics during early double-strand break processing revealed by non-intrusive imaging of living cells. *PLoS Genet.* **10,** e1004187 (2014).
214. Dostie, J. & Dekker, J. Mapping networks of physical interactions between genomic elements using 5C technology. *Nature Protocols* **2,** 988–1002 (2007).
215. Boettiger, A. N. *et al.* Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature* **529,** 418–422 (2016).
216. Lindhout, B. I. *et al.* Live cell imaging of repetitive DNA sequences via GFP-tagged polydactyl zinc finger proteins. *Nucleic Acids Research* **35,** e107–e107 (2007).
217. Ma, H., Reyes-Gutierrez, P. & Pederson, T. Visualization of repetitive DNA sequences in human chromosomes with transcription activator-like effectors. *Proceedings of the National Academy of Sciences* **110,** 21048–21053 (2013).
218. Miyanari, Y., Ziegler-Birling, C. & Torres-Padilla, M.-E. Live visualization of chromatin dynamics with fluorescent TALEs. *Nat. Struct. Mol. Biol.* **20,** 1321–1324 (2013).
219. Thanisch, K. *et al.* Targeting and tracing of specific DNA sequences with dTALEs in living cells. *Nucleic Acids Research* **42,** e38–e38 (2014).
220. Chen, B. *et al.* Dynamic Imaging of Genomic Loci in Living Human Cells by an Optimized CRISPR/Cas System. *Cell* **155,** 1479–1491 (2013).
221. Anton, T., Bultmann, S., Leonhardt, H. & Markaki, Y. Visualization of specific DNA sequences in living mouse embryonic stem cells with a programmable fluorescent CRISPR/Cas system. *Nucleus* **5,** 163–172 (2014).
222. Chen, B. *et al.* Expanding the CRISPR imaging toolset with Staphylococcus aureusCas9 for simultaneous imaging of multiple genomic loci. *Nucleic Acids Research* gkv1533–13 (2016). doi:10.1093/nar/gkv1533
223. Ma, H. *et al.* Multicolor CRISPR labeling of chromosomal loci in human cells. *Proc Natl Acad Sci USA* **112,** 3002–3007 (2015).
224. Wang, S., Su, J.-H., Zhang, F. & Zhuang, X. An RNA-aptamer-based two-color CRISPR labeling system. *Sci Rep* **6,** 26857 (2016).
225. Tu, L.-C. *et al.* Multiplexed labeling of genomic loci with dCas9 and engineered sgRNAs using CRISPRainbow. *Nat. Biotechnol.* 1–4 (2016). doi:10.1038/nbt.3526
226. Carlson, D. F., Fahrenkrug, S. C. & Hackett, P. B. Targeting DNA With Fingers and TALENs. *Molecular Therapy - Nucleic Acids* **1,** e3–4 (2012).
227. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas

systems. *Science* **339,** 819–823 (2013).
228. Dominguez, A. A., Lim, W. A. & Qi, L. S. Beyond editing: repurposing CRISPR–Cas9 for precision genome regulation and interrogation. *Nature Publishing Group* **17,** 5–15 (2015).
229. Hess, G. T. *et al.* Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells. *Nat Meth* **13,** 1036–1042 (2016).
230. Qi, L. S. *et al.* Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell* **152,** 1173–1183 (2013).
231. Li, X.-T. *et al.* tCRISPRi: tunable and reversible, one-step control of gene expression. *Sci Rep* 1–12 (2016). doi:10.1038/srep39076
232. Chen, B., Guan, J. & Huang, B. Imaging Specific Genomic DNA in Living Cells. *Annu Rev Biophys* **45,** annurev–biophys–062215–010830–23 (2015).
233. Nishimasu, H. *et al.* Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA. *Cell* **156,** 935–949 (2014).
234. Terns, R. M. & Terns, M. P. CRISPR-based technologies: prokaryotic defense weapons repurposed. *TRENDS in Genetics* **30,** 111–118 (2014).
235. Bates, M., Dempsey, G. T., Chen, K. H. & Zhuang, X. Multicolor super-resolution fluorescence imaging via multi-parameter fluorophore detection. *Chemphyschem* **13,** 99–107 (2012).
236. Lubeck, E. & Cai, L. Single-cell systems biology by super-resolution imaging and combinatorial labeling. *Nat Meth* **9,** 743–748 (2012).
237. Singh, D., Sternberg, S. H., Fei, J., Doudna, J. A. & Ha, T. Real-time observation of DNA recognition and rejection by the RNA-guided endonuclease Cas9. *Nature Communications* **7,** 1–8 (2016).
238. Hendel, A. *et al.* Chemically modified guide RNAs enhance CRISPR-Cas genome editing in human primary cells. *Nat. Biotechnol.* **33,** 985–989 (2015).
239. Zetsche, B. *et al.* Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell* **163,** 759–771 (2015).
240. Yamano, T. *et al.* Crystal Structure of Cpf1 in Complex with Guide RNA and Target DNA. *Cell* 1–15 (2016). doi:10.1016/j.cell.2016.04.003
241. Tsai, S. Q. *et al.* Genome-wide specificities of CRISPR-Cas Cpf1 nucleases in human cells. *Nat. Biotechnol.* **34,** 882–887 (2016).
242. Nilsen, T. W. 3'-End Labeling of RNA with [5'-32P]Cytidine 3‘,5’-Bis(Phosphate) and T4 RNA Ligase 1. *Cold Spring Harbor Protocols* **2014,** pdb.prot080713–pdb.prot080713 (2014).

Curriculum Vitae

Xiaoli Weng Updated: Nov 5, 2017
xiaoliweng@gmail.com
Place of birth: Suzhou, China
Citizenship: US

Academic Background

2017 Ph. D. Program in Molecular Biophysics, Johns Hopkins School of Medicine

2010 M.A. Chemistry, Wesleyan University

2009 B.A. Molecular Biology and Biochemistry, Wesleyan University

---

Research Experience

Johns Hopkins University School of Medicine, Baltimore, MD
2011-present
Graduate Researcher, Laboratory of Professor Jie Xiao
Investigation of the spatial organization of transcription in *E. coli* using superresolution imaging microscopy

- Probed the spatial organization of molecular components of transcription such as RNA polymerase, chromosomal DNA and RNA transcripts using techniques including PALM/STORM, smFISH.

Experimentally measured transcription factor mediated chromosomal DNA loop formation inside live *E. coli* using a modified Fluorescent Reporter Operator System

- Using 2-color imaging, measured and quantified the distance separation between two target DNA locations as close as ~2.3 kb apart linearly, and correlated spatial distance separation with DNA loop formation.

Development of *in vivo* DNA imaging via CRISPR Cas9 system

- Synthesized various RNA substrates for CRISPR Cas9 DNA targeting via in vitro transcription and enzymatic dye labeling. Performed in vitro cleavage assays to assess targeting efficiency of sgRNA sequences.

Wesleyan University, Middletown, CT
2007-2010
Undergraduate/Graduate Researcher (BA/MA), Laboratory of Professor Irina Russu
Investigation of the dynamics and energetic signatures of the tR2 intrinsic terminator of the bacteriophage lambda

- NMR experiments were performed to measure the exchange of imino protons with solvent protons, this exchange rate is used to calculate the rate of the opening and closing of individual base pairs within a short nucleic acid duplex.

Teaching and Service

2013-2015 Volunteer Instructor, JHMI Fun with Science Summer Camp (for young students)
2013-2015 Volunteer Instructor, JHMI Science Day (for young students)

2009 Teaching Assistant, General Chemistry, Wesleyan University

Awards and Honors

2017 Biophysical Society Thematic Meeting (Taipei) Travel Award

2008 HHMI summer research fellow, Wesleyan University

Publications

Weng X, Xiao J. (2014) Spatial organization of transcription in bacterial cells. *Trends Genet.*

Hensel Z, **Weng X**, Lagda AC, Xiao J. (2013) Transcription-factor-mediated DNA looping probed by high-resolution, single-molecule imaging in live *E. coli* cells. *PloS Biol.*

Huang Y, **Weng X**, Russu IM. (2011) Enhanced base-pair opening in the adenine tract of a RNA double helix. *Biochemistry*

Huang Y, **Weng X**, Russu IM. (2010) Structural energetics of the adenine tract from an intrinsic transcription terminator. *J Mol Biol.*

Oral Presentations

Nov 2017 *Spatial organization of transcription in E. coli*. Chesapeake Bay Single Molecule Conference, Potomac, MD

Jun 2017 *The spatial and functional organization of transcription in E. coli*. Biophysical Society Thematic Meeting - Single-cell Biophysics, Taipei, Taiwan

Mar 2017	*The spatial and functional organization of transcription in E. coli*. JHMI Center for Cell Dynamics Seminar Series, Baltimore, MD
Sept 2016	*The spatial and functional organization of transcription in E. coli*. Institute for Biophysical Research Retreat, Baltimore, MD
May 2016	*Single molecule imaging of bacterial transcription.* JHMI Topics in Macromolecular Structure and Function Seminar Series, Baltimore, MD
May 2013	*Molecular diffusion in cellular processes*. JHMI Center for Cell Dynamics Seminar Series, Baltimore, MD

Poster Presentations

Jun 2017	*Spatial organization of transcription in E. coli cells via superresolution fluorescence microscopy.* Biophysical Society Thematic Meeting - Single-cell Biophysics, Taipei, Taiwan
Apr 2017	*Spatial organization of transcription in E. coli cells via superresolution fluorescence microscopy.* Chesapeake Bay Single Molecule Conference, Baltimore, MD
Feb 2017	*Spatial organization of transcription in E. coli cells via superresolution fluorescence microscopy.* Biophysical Society Meeting (61st), New Orleans, LA
Feb 2015	*Spatial organization of transcription in E. coli cells via superresolution fluorescence microscopy.* Biophysical Society Meeting (59th), Baltimore, MD
Jun 2013	*Spatial organization of transcription in E. coli cells via superresolution fluorescence microscopy.* FASEB Prokaryotic Transcription Conference, Saxtons River, VT
Mar 2011	*Spatial organization of transcription in E. coli cells via superresolution fluorescence microscopy*. Biophysical Society Meeting (55th), Baltimore, MD